UNIVERSITY OF CALGARY

The Feasibility and Fidelity of Practicing Surgical Fixation of an Ulna Fracture on Virtual Bone

by

Justin LeBlanc

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL

FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MEDICAL SCIENCE

CALGARY, ALBERTA

OCTOBER, 2009

ABSTRACT

The goals of this study were to: 1) evaluate the fidelity, and 2) determine the feasibility of developing a high fidelity virtual surgical simulator. A stratified, randomized, within-subjects design compared surgical fixation of the ulna using the Sawbones and newly created virtual simulators. Participants were assessed using itemized checklists, global rating scales (GRS), fidelity questionnaires, and measures of pre-/post-procedure knowledge, skill and comfort levels for surgical fixation of the ulna. Construct validity was demonstrated for both simulators ($p<0.05$). Reliability of the combined checklist and GRS were $\alpha > 0.8$, while the intraclass correlation coefficients were > 0.9. Fidelity of the virtual simulator was rated lower than the Sawbones simulator ($p<0.001$). Costs of a virtual simulator are higher initially than the Sawbones simulator, but these costs are similar after 10 years. Although the virtual simulator demonstrated construct validity, a high fidelity surgical fixation virtual simulator requires greater time and financial resources.

ACKNOWLEDGEMENTS

I especially want to thank the following persons for their encouragement and support during my Masters studies and the completion of my thesis:

- Dr Tyrone Donnon, my excellent supervisor. Thank you for all your time, effort, and patience in helping me meet all my deadlines, and for ultimately getting me through my Master's. I look forward to working with you in the future.

- Dr Carol Hutchison. You started me on this path to a medical education degree and I would like to thank you for your great ideas, support and help while completing this project.

- Dr Yaoping Hu. Thanks to you and the Department of Electrical and Computer Engineering for volunteering all of your time, rooms and equipment, without which this project would have never been able to be such a success.

- Manar Din Samad, Aron Su and Antoine Widmer, who with the help of Dr Hu, developed the virtual ulna simulator that was used for this project.

- Dr Tanya Beran, who was finally able to help me understand not just how to properly use statistics, to why I am using it. Thanks again for all your help and support.

- Dr Gail Kopp, for starting me off with a quest to determine exactly what fidelity is, and for supporting me in the completion of this project, thank you.

- To my examiners who volunteered their time so graciously, Dr Jeremy Tillim, Dr Ganesh Swamy, Dr Beth Pedersen and Dr Tom Van Raaij (Netherlands).

- To all of the hard working residents from the University of Calgary Division of Orthopaedic Surgery. Your 100% volunteer rate is unheard of in any study and your participation was much appreciated.

- To Dr Jacques Bouchard, Dr Simon Goldstein and all the staff from the Division of Orthopaedic Surgery, who provided me with their time and support, allowing me to complete this degree during my residency.

- To the Department of Surgery and especially the Surgical Scientist program, which funded me through this year and enabled me to take time off and complete this degree.

- To Synthes Canada, and Archie Yamada, for donating the surgical tools, Sawbones and time, allowing the completion of this project

- And last, but definitely no least, to my beautiful fiancé Carmen Malenica, who supported me throughout the past years and read through most of this thesis without once complaining, your are the best, thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES AND ILLUSTRATIONS

LIST OF SYMBOLS, ABBREVIATIONS AND NOMENCLATURE

Symbol	Definition
OSATS	**Objective Structured Assessment of Technical Skills**
NASA	**National Aeronautics and Space Administration**
MCC	**Medical Council of Canada**
US	**United States of America**
ACLS	**Advanced Cardiac Life Support**
ICU	**Intensive Care Unit**
MIST-VR	**Minimally Invasive Surgical Trainer - Virtual Reality**
GRS	**Global Rating Scale**
MANOVA	**Multivariate analysis of variance**
ICSAD	**Imperial College Surgical Assessment Device**
U of C	**University of Calgary**
PGY	**Post-Graduate Year (1,2,3,4,5)**
AO	**Arbeitsgemeinschaft für Osteosynthesefragen**
AHS	**Alberta Health Services**
SPSS	**Statistical Package for the Social Sciences**
d	**Cohen's *d* = effect size difference**
ANOVA	**Analysis of variance**
SES	**Standard error of skewness**
SEK	**Standard error of kurtosis**

r	**Pearson product moment correlation**
α	**Cronbach's α = internal consistency**
κ	**Cohen's Kappa = inter-rater reliability**
ID	**Identification**
M	**Mean**
SD	**Standard deviation**
pre	**Pre-procedure questionnaire**
post	**Post-procedure questionnaire**
VR, vr	**Virtual reality simulator**
Saw, S	**Sawbones simulator**
sec	**Seconds**
Enviro	**Environment domain**
Equip	**Equipment domain**
Psych	**Psychological domain**
ICC	**Intraclass coefficient**

CHAPTER 1 – INTRODUCTION

Background

Surgical resident education is based on a foundation of clinical experiences, didactic teaching and studying. Surgical skills are most frequently acquired through graduated responsibility in the operating room under the supervision of a practicing surgeon. This system has proved to be effective, however it is not without risks, the first of which is patient safety. Surgical techniques in the hands of the inexperienced may bring injury to the patient. Second is the cost to the system of training these surgeons, which results in longer operating times and higher operating costs [1,2]. It is expected that the more experience a resident has prior to entering the operating room, the more benefits in terms of time, cost and safety can be realized. For these reasons there has been increasing interest in practicing with hands-on models, cadavers [3-5], and recently the use of virtual reality simulators [1,3,6-8].

Ethical, economic and educational considerations have led to development of alternative methods of teaching and training in surgical techniques [2]. Cadavers can be used for the training of new techniques and are beneficial to the surgical trainee due to their realism. The use of cadavers for purposes of surgical training is limited, however, by supply, cost and consistency. Synthetic plastic bone (Sawbones as one example) is also used for training a wide variety of orthopaedic surgical skills. There are many types of synthetic bone that differ in cost and quality, all of which attempt to simulate real bone as closely as possible. The typical use of Sawbones includes anatomical reduction of the fracture followed by internal fixation using specialized implants and techniques.

The idea of utilizing computers for surgical training brings forth the advantages of repeated use without added cost, possibilities for immediate feedback and opportunities to input potential surgical errors into the simulation that would provide the residents with experience in handling problems with these procedures without risking a patients safety. Advances in tissue modeling, graphics and haptic (force-feedback) instrumentation have led to the development of useful virtual reality machines [1,6,7,9-11]. Validation studies using these simulators have shown differences between novice and experienced surgeons and improvement in training scores over time, and simulator task performance can be correlated to actual task performance [1,2,12-14]. Ideally, training methods should be used to reduce operating time and errors, as well as promote confidence and competence in surgical skills by increasing resident knowledge, skills and comforts.

High quality simulators were first used for training civilian, military pilots and astronauts [1]. The use of simulators is now present within the engineering, architecture and entertainment industries [15]. A very promising area for this technology is medicine. Currently the use of simulation technology can be found in endoscopic, laparoscopic, vascular, dermatological and neurosurgical procedure training [1,3,7,8,10,12]. These simulators provide surgeons with the opportunity to practice with virtual body tissue and receive feedback, similar to that of a real operation [15]. These simulators have been shown to measure and improve surgical skills [9,12,14].

A task force for evaluating virtual reality in orthopaedic surgery was formed in 1998 [6,16]. They proposed to identify the application of virtual reality in an orthopaedic surgery sense, establishing guidelines to develop virtual reality simulator programs and evaluating virtual reality technology. Virtual reality arthroscopic machines have now been developed and are in the process of being validated for knees [6,16,17] and shoulders [1,18-20].

The transfer of skills learned on virtual reality simulators has been encouraging, demonstrating that the trainees performed the dissections more quickly, made fewer errors and had higher economy of motion scores than did those trainees without such training [2,21]. Soft tissues, bones and the use of cutting and retracting tools have all been successfully implemented into virtual reality simulators [7,8,10,11]. There has, however, been no development of simulators that allow residents to practice the surgical fixation of common orthopaedic fractures. Our interest lie in whether a virtual reality simulator can be used to aid in teaching residents surgical fixation of the ulna. The feasibility and fidelity of this simulator will be assessed and compared to the use of Sawbones for fracture fixation.

Educational Significance

New teaching techniques must be developed so that the knowledge, skills and comforts of residents, and the safety of the patients will be apparent in the operating room. At this time, Sawbones and cadavers are more commonly used for surgical training, but bring with them a great expense. A new method of learning is needed and virtual simulation has proved to be a valuable tool in many surgical fields. In orthopaedics, only arthroscopic virtual reality devices have been developed thus far. It is felt that virtual simulators can be used for training in the field of orthopaedic trauma. To assess this, surgical fixation of a fractured ulna can be performed with a virtual simulator and compared to that of the current method of training, a Sawbones simulator. They will be evaluated using newly developed questionnaires, task-specific checklists and global rating scales. If this virtual surgical simulator and the evaluation tools are proven to be valuable, they may then be further used to evaluate more complicated fracture fixation techniques. This simulator may be useful for surgical residents' hands-on learning of fracture fixation skills. It may also become a useful evaluation tool for programs to ensure proper techniques are being used.

Statement of problem

The research questions of this study were:

1) Does surgical fixation of an ulna fracture on a virtual bone simulator have the same or comparable fidelity at all levels of resident training as performing the same subtasks on Sawbones?
2) Can we measure participant performance and quality of surgical skills using a modified checklist and global rating scale format?
3) Can we accurately assess the learners perceptions of their experiences related to specific subtasks of the procedure using a questionnaire;
4) Is it feasible to develop a virtual fracture fixation model for orthopaedic surgery residents to perform surgical fixation of an ulna?

We hypothesize that the educational benefits and fidelity of this model will be equivalent to a Sawbones model at all levels of resident training. We also hypothesize that a feasible virtual simulation model with force (haptic) feedback will be able to be developed for the training of orthopaedic surgery residents in fracture fixation and will be practical for the training of orthopaedic surgery residents in the future.

Chapter 2 contains a detailed review of the current literature concerning virtual simulators and certain assessment tools. Chapter 3 will describe the simulators, how the experiment was carried out and how the data was analyzed. Chapter 4 presents the findings of the simulators and participants. Chapter 5 summarizes the findings and provides a discussion of the results in regards to surgical education.

CHAPTER 2 – LITERATURE REVIEW

This chapter contains a review of current literature aimed at increasing one's understanding of using virtual simulation for surgical education. The discussion will include the educational theory of psychomotor skills, an understanding of what the fidelity of a simulator is, the history of medical simulators and the validation of these simulators, and the review of the current state of virtual orthopaedic simulators. The following literary review is intended to develop an appreciation for simulation in surgical education as a valuable tool. In addition, and of noteworthy importance, is how these simulators can be used in the assessment of psychomotor skills, using both a currently accepted model of the objective structured assessment of technical skills (OSATS) and questionnaires.

Virtual Reality in Surgical Education

In 1970, the NASA ground crew was able to aid Apollo 13 in safely returning to earth with the use of a flight simulator, and since that time the airline industry has been using advanced simulators to prevent catastrophes [22]. A simulator can be defined as a device that allows a participant to reproduce phenomena under test conditions that mimic actual environments, with sufficient realism to eliminate participant disbelief [23]. Since 1955, the Federal Aviation Administration has used simulation in the recertification of commercial pilots' licenses [24]. The development of digital computers has facilitated the creation of new flight simulators with increasing realism. These simulators have enabled pilots to perform textbook takeoffs and landings without stepping a foot in an airplane. As the aviation industry uses simulation to provide pilots with a safe arena to acquire the necessary skills to safely manage an airplane, the medical community should consider the development of simulators for the safety of the medical staff and patients alike. Compared to the aviation industry, simulation in the medical field is still in its infancy. However, prior to discussing how simulation is currently utilized in medical education, discussions of both the educational theory for psychomotor skill acquisition and of fidelity are required.

Educational Theory

Simulation use in medical education stems from the experiential educational theory proposed by Carl Rogers. This is an adult learning theory in which Rogers refers to experiential learning as applied learning [25]. This type of learning address' the needs and wants of the learner, is self-directed and is evaluated by the learner. A learning environment is thought to be at its best when external and self-threats are minimized and when the learner believes the subject matter is relevant. A teacher can best facilitate this learning when a positive learning environment is set, resources are organized and easily available, and when objectives are clarified [25]. Simulation also follows the learning theories of: behaviourism: students practice until properly trained; cognitivsm: learners acquire and reorganize knowledge through the use of simulation; and constructivism: as the student builds upon an acquired knowledge base by interacting with the learning material.

Learning can be broken down into three main domains: cognitive (knowledge), affective (attitude/comfort) and psychomotor (skills). These domains were first introduced to aid teachers, professional specialists and research workers who were dealing with curricular and evaluation problems, and to suggest the types of objectives to be included in their curricula. The cognitive domain provided a classification of educational goals dealing with the recognition of knowledge and the development of intellectual skills and abilities [26]. The affective domain deals with appreciations, attitudes, desires and interests [27]. The cognitive and affective domains were subdivided, starting from the simplest behaviour to the most complex. The same group who proposed the first two domains recognized the existence of a third domain – the psychomotor domain. They did not, however, develop a classification system for it as they felt they had little experience in teaching manual skills at the secondary and college levels, and did not believe in its development.

An important domain, more so in the surgical specialties than other medical specialties, is the psychomotor domain. Psychomotor skills can be taught in both the operating room and in a skills laboratory. Educators eventually felt a need for a classification system for the psychomotor domain and for its use in curricula as well as for a basis of evaluating education. They felt it was needed for education in general, but also in specialized areas such as industrial education, agriculture, music, art and physical education [28]. The classification for the psychomotor domain is useful for research and in the teaching and development of motor abilities and skills. Seven major subdivisions can be described similar to the cognitive and affective domains, from the simplest to the most complex: Perception, Set, Guided Response, Mechanism, Complex Overt Response, Adaptation and Origination.

Perception, the simplest level, is the process of becoming aware of objects and using sensory cues to guide motor activities, such as estimating where a thrown ball will land. **Set** is the readiness to act in the mental, physical and emotional sets, such as knowing your own limitations of a task. **Guided response** is knowing the required steps needed to complete a task or skill, such as following instructions to build a model. **Mechanism** is an intermediate step in learning complex skills. The person can perform the skill in a somewhat confident and proficient manner, such as driving a car. **Complex overt response** is when the person can skillfully perform a motor act involving complex motion patterns, such as maneuvering a car into a tight parking spot. **Adaptation** is when skills are well developed, but the person can also modify their actions to account for new situations, such as driving a car when the tire pops. **Origination** is the ability to create a new task or skill to fit a specific situation, incorporating learned tasks. This emphasizes creativity from highly developed skills, such as creating new training programs.

Objective structured assessment tools used to assess surgical skills may be used in laboratory settings to effectively evaluate surgical trainees during their residency and to ensure appropriate technical skills are being acquired to develop competent surgeons. An effective assessment tool for surgical trainees should aim to evaluate from the perception stage up to and including the complex overt response. At times, it may be possible to reach the adaptation stage with such an assessment or evaluation tool.

Minimally invasive procedures require increased skills and are associated with long learning curves [29]. The development of laparoscopic and arthroscopic skills initially involves learning both cognitive and psychomotor skills, and then refining them [30]. It has been suggested that early training outside the operating room could result in safer, more efficient and cheaper training than in the operating room [31]. Performance curves are perhaps a better name for these curves as learning cannot actually be measured when looking at these curves [32]. These performance curves are evaluated in surgical education and it has been shown that a novice may reach the plateau of the curve (full effect of learning perceived) in as few as 5 repeated attempts of the task on a virtual surgical simulator, for time to completion, total path length, number of movements and camera navigation [9,29]. For more advance tasks, the novice may require a much higher number of attempts, such as for lifting and grasping [9], and as high as fifty-three times for a novice to achieve 90% proficiency in drilling bone [33]. It is suggested that surgical errors occur more consistently in the beginning (steepest) portion of the performance curve, where the improvement rate is also the highest [34]. A performance curve can be utilized to assess how many repetitions of a task must be completed to progress past the steep phase of the curve, and haptics (force-feedback) has been suggested to aid in decreasing this steep portion of the performance phase [35].

Fidelity in Medical Simulation

Fidelity can be defined as the extent to which the appearance and the behaviour of the simulation match that of the real environment [36]. Fidelity can be further divided into physical and psychological fidelity.

Physical fidelity is the degree to which the device or environment actually replicates the physical characteristics of the real task, or the degree to which the simulation looks, sounds, and feels like the operational environment in terms of visual displays, controls and audio [37]. Haptics is included in the physical fidelity domain, and can be defined as the manual interaction in which the participant can touch, feel, and manipulate objects in the simulated environment [38]. Haptics is used to provide the sense of resistance that would normally be felt in the real situation as objects come into contact with each other. It has been suggested that haptics will increase the fidelity of a simulator [39]. Physical fidelity can be further divided into environmental and equipment domains. Environment domain is the extent to which the simulator duplicates motion, visual and other sensory information from the true environment. Equipment domain is the degree to which the simulator duplicates the appearance and feel of the real system.

Perhaps of greater importance is psychological fidelity, which can be described as the degree to which simulation replicates psychological factors, such as stress and fear, which can be experienced in the real environment [37]. Higher levels of psychological fidelity in a simulator may be associated with higher degrees of skill or knowledge transfer [37]. Transfer is the transmission of knowledge, skills and comforts developed during training, which are transmitted to the real environment in which they are normally used. A simulator should aim to provide positive transfer, in which learning in the training environment improves performance in the targeted environment [37]. Negative transfer should be avoided, where learning in the training environment worsens performance in the targeted environment. An example of negative transfer is when surgical trainees practice on a simulator in which poor techniques may be acquired. Negative transfer may result in poor performance and compromised patient safety.

Simulators can be defined as either high or low fidelity. High fidelity simulators offer more real-life qualities, which immerse learners in a more realistic interactive environment. Low fidelity simulators use materials and equipment that are less similar than what is used in the true environment, and may be prone to negative transfer. It has been argued that basic skills can be learned well on a low fidelity model, whereas more technical skills require higher fidelity models [37,40]. An example of high versus low fidelity simulators is to compare training for microsurgery on either live rat vas deferens (high fidelity) or silicone tubes (low fidelity). A study compared novice surgical trainees using these two training techniques. It was found that to perform basic surgical skills (suturing), there was no significant difference between high and low fidelity simulators for the transfer of basic skills, demonstrating that low fidelity models may be sufficient for junior surgical trainees [41]. Similar results were found with surgical trainees learning basic skills on low fidelity laboratory bench simulators. The trainees were able to transfer these learned skills onto a cadaver [42]. Junior surgical trainees may therefore begin training on low fidelity simulators to learn basic techniques, and then graduate to higher fidelity simulators in order to acquire higher levels of surgical skills.

Simulation in Medical Education

Accepting the use of simulation in medical education has taken much longer than it has in the aviation industry, but basic models have been utilized for learning pathology and anatomy for centuries [43]. Simulator use in medical education can be divided into five categories: Verbal, Standardized patients, Part Task Trainers, Computer Patient and Electronic Patient [44]. Verbal simulation can best be thought of in the terms of role-playing. This type of simulation is useful for medical education at all levels, and can be used in medical schools for practicing and developing history taking and physical exam skills.

Standardized patients were first used in 1963, with the intent of helping third year medical students prepare for their neurology rotation [45]. These simulators are actors who are taught to portray different and difficult patient experiences and have been brought into use in many aspects of medical education. Since 1993, the use of standardized patients have been included in the licensure examinations of the Medical Council of Canada (MCC), and since 2004 for the Step II Clinical Skills for US medical students [24].

Part-task trainers are anatomic models of body parts, either visualized in a normal or diseased state. The first such model was designed by Åsmund Lærdal, a Norwegian toy manufacturer. With the help of anesthetists, Lærdal developed the “Resusci-Anne”, the part task trainer for resuscitation training [43]. Since that time, many new higher fidelity models have been developed. The Visible Human project, which began in 1994, has aided greatly in supplying the correct anatomical visualization of these models [46]. Task trainers can now be seen in a variety of specialties, and some will be discussed later.

The computer patient is an interactive software or Internet based virtual simulator. The first of these simulators was a resuscitation model developed in 1983 [24]. This theme was built upon, and by 1995 the Anesthesia Simulator Consultant was created [47]. This model has led to multiple anesthesia, ACLS and cardiology interactive and web based training simulators, all of which are frequently used for their respective specialty trainees.

Electronic patients are the latest to be developed. These are mannequins or virtual reality based simulators that replicate the clinical environment and the patient. The first full scale human patient mannequin was developed by a group of anesthetists in the late 1980’s [48], and since then many high fidelity full body simulators have been developed for anesthesia, ICU, trauma, and emergency medicine for both adult and paediatric medicine [24,49]. Virtual reality patients can either be whole body or used as part-task trainers, and these will be discussed in the following section.

Virtual Simulation

Virtual reality refers to “the recreation of environment or objects as a complex, computer generated image” [43]. The main aim of this type of simulator is to present virtual objects or environments to all human senses, in the same way they would appear in their natural environment. Initial virtual simulators were of lower fidelity due to the lack of force feedback and technology. The newer simulators contain haptics, which as discussed prior replicate the kinaesthetic and tactile perception of the real experience. This helps to produce the sensation of resistance when using instruments within the simulated environment, in which the learner feels as though they are coming into physical contact with the simulated object.

The first virtual simulator was created in 1961 by Morton Heilig, in which the user would insert 25 cents, sit inside a booth containing a movable seat, and enjoy a ten-minute virtual experience. It was called the Sensorama, which projected three dimensional stereoscopic images, vibrations with body tilting, stereo sound, aromas and wind in five different immersive experiences, one of which was a bicycle ride through the streets of Brooklyn [24]. Virtual reality developed further during the 1980's, bringing forth head mounted displays, full body suits and eventually haptics [13]. The advancement of computer graphics throughout the years has been greatly aided by the gaming industry, helping to increase the physical fidelity of virtual simulators.

Richard Satava, a professor of Surgery and Senior Science Advisor at the US Army Medical Research in Maryland, was the surgeon on the team that developed the first surgical robot and helped develop one of the first surgical simulators. Dr. Satava has found many advantages with virtual reality simulators and has stated:

> "The greatest power of virtual reality is the ability to try and fail without consequence to animal or patient. It is only through failure, and learning the cause of failure, that the true pathway to success lies" [50]

Virtual simulation is a valuable innovation as it creates a non-threatening, risk-free training environment in which the learner can repeatedly practice specific skills. It can provide realistic scenarios with unbiased assessment and immediate objective, non-biased feedback. With the increasing costs of training surgical residents – operating room costs alone are over $48,000 to train a general surgery resident for four years [51] – virtual simulation is a new alternative that may lead to a decrease in the costs of training in the operating room. Virtual simulation is a new educational method that may be utilized in many surgical specialties. However, prior to incorporating a virtual reality simulator into a surgical training curriculum, it must be proven to aid a learner in acquiring basic technical skills, improve the learners' performance in an operating room and be verified as a valid simulator for surgical procedures.

Virtual surgical simulators

Virtual surgical simulators were first introduced in the 1990's. The first was a lower extremity tendon transfer simulator, which computed the force and movement of each musculo-tendinous complex when surgically transferred in the lower limb, allowing a surgeon to practice tendon transfers and lengthening in a non-threatening environment [52]. The next was an open abdominal simulator for exploration of the organs using basic surgical tools, used mostly for anatomic education [53]. Neither of these early simulators were validated nor did they demonstrate the acquisition of technical skills, but many studies have been able to do so since. Most laparoscopic studies have been completed using the Minimally Invasive Surgical Trainer - Virtual Reality (MIST-VR). This laparoscopic virtual reality trainer has been employed since 1998, and novices have demonstrated acquiring skills to the same degree that they do on conventional video box trainers (gold standard), significantly improving their skills compared to those with no training [54-56]. These studies demonstrate that through virtual simulator use, a trainee can acquire surgical skills. They did not, however, demonstrate a transfer of skills to real patients, nor did they show a difference between the current methods of training (video box trainer) and the virtual reality models.

Grantcharov (2004) performed a randomized controlled trial using 16 surgical residents and compared MIST-VR training to no training. These residents were then evaluated while performing a human laparoscopic cholecystectomy [2]. The MIST-VR group demonstrated shorter times, fewer errors and better economy of moves, demonstrating a transfer of skills from the virtual trainer to real life. In 2002, Hamilton demonstrated again that the MIST-VR improved operative performance [57]. Fifty surgical residents were randomized to either train using the video box trainer or virtual reality simulator (MIST-VR), and were assessed during a human laparoscopic cholecystectomy. A significant improvement from baseline for the virtual reality group was noted, whereas no significant improvement could be demonstrated for the box trainer group. When directly comparing these groups, no significant differences of their final skills scores or their global assessments for human laparoscopic cholecystectomies were noted. These studies demonstrate that surgical trainees can learn psychomotor skills from a virtual reality simulator (MIST-VR), and that a transfer of skills to human patients can occur, but they do not demonstrate that virtual simulators are better than the current gold standards (video box trainer for laparoscopic skills).

Preceding the use of a virtual simulator in a surgical training curriculum, it must be validated. A valid virtual simulator should be able to imitate the visual and spatial environment, the real time characteristics of the simulated procedure, and if possible include a haptics (force-feedback) device [58]. A simulator or evaluatory tool cannot be proven valid within a single experiment, but needs to be proven over time as it accrues evidence to its validity. Construct validity is often used as an alternative to measure validity in the case of a single experiment. For measuring validity in surgical skills, construct validity implies that individuals who are more likely to perform better in a real task should perform better on the simulated task. Before suggesting a simulator is valid, it should provide not only construct validity, but also content and/or criterion validity (as measured against a gold standard simulator).

Content Validity:

The first validity to be determined should be content validity. This is a non-statistical, expert and judgment based validity. It is based on the description of the contents of the simulator, and judgment about to what extent the simulator covers the subject matter of the procedure in real life [30]. For example, how much a virtual simulator designed to develop the skill of drilling bone actually measures the true psychomotor skill of drilling bone in real life. Content validity is not often mentioned, as it is essentially a decision made early in the development of a simulator by the experts involved with its creation, and there is no statistical method or means to prove it. However, some articles rely on face validity as a criterion to help explain content validity.

Face Validity:

Face validity refers to whether a model actually resembles the true task it is based on, or refers to the degree to which a test appears to measure what it claims to measure [59]. In other words, it reflects the qualitative user perceptions of the simulator [17]. For studies providing face validity, a questionnaire is often used to determine if the participants in the study believe the instrument simulates what it is supposed to, and whether it would be a useful tool for training [30,60]. It is important to get the opinions of both the expert and novice trainees as it is the novices who will be using these simulators to train. The use of this type of validity alone is weak, so it should be used in conjunction with others.

Criterion Validity:

One form of criterion-related validity is concurrent validity, which refers to the degree that a test score on a simulator compares to that of another previously validated instrument that supposedly measures the same construct [59]. Concurrent validity is measured by determining the correlation of scores between the new simulator and those of the previously established measuring instrument. The participants complete both tests and then a relationship between the two tests are determined.

Predictive validity is another form of criterion validity. Predictive validity refers to whether a test can predict how well an individual will do in a future situation [59]. For example, the results of a simulator test at the start of the year will be able to predict which residents will do well or poorly at the end of the year on that same test.

Construct Validity:

Construct validity is the most important form of validity as it refers to what the test actually measures. It evaluates the simulator based on the degree to which it identifies the quality, ability or trait it was designed to measure [30].

The current method most often used in laparoscopic procedures for validating a simulator is to measure the inverse transfer of skills. These experiments compare the performance of an expert surgeon with that of a novice on the simulator, where the experts should perform better on the simulator than the novices [58,61]. Some laparoscopic simulators have used this method alone to suggest that this type of validity is enough to determine that it is a useful tool for surgical education [62]. This method should not be used alone as it does not allow one to accurately conclude that the specific training completed on a surgical simulator promotes positive transfer of skills [63].

Using the MIST-VR trainer, Gallagher (2003) assessed if virtual simulator performance correlated with surgical experience [61]. Indeed this trainer could distinguish between the novice (medical student with no laparoscopic experience) and an expert (>50 procedures completed). Other studies have also demonstrated construct validity for virtual simulators [30,31,62,64]. These studies demonstrate the construct, face and concurrent validity for laparoscopic virtual simulators, but still leaves the question of whether virtual reality simulators are better than the video box trainers unanswered.

Most researchers seem to agree that surgical virtual simulators are useful tools for learning surgical skills. With the increasing costs of training a resident in terms of operating time, the development of minimally invasive procedures which require more technical skills, the shortening of a resident's work week, and the increasing pressure for safe, ethical practice, training methods must be developed to train surgical residents outside the operating room. The unanswered question at this time is which is better - the traditional low fidelity training methods (video box trainer) in which you can use real tools and modified anatomical structures, or high fidelity virtual reality simulators in which you can practice multiple times, create new scenarios and obtain immediate feedback. There is no evidence than one training method is better than the other, so a cost-benefit analysis could be conducted to determine which is more beneficial. Many authors comment that virtual reality may actually decrease training costs overall in the long term, as cadavers and materials for video box trainers are often costly themselves. However, literature is lacking at this time to help determine if in actuality a virtual reality simulator is less expensive overall than the lower fidelity models currently being used.

It is known that both video box trainers and virtual reality simulators are effective learning methods for surgical skills. So with no cost-benefit analyses currently available, how can a surgical training program choose which method is superior? One approach is to ask surgical trainees. In 2002, Hamilton asked the participants which method of training they preferred [57]. Eighty-three percent of the residents thought the video box trainers were more effective than the virtual simulator (MIST-VR) and seventy-seven percent preferred the box trainer to the virtual simulator. They felt the box trainer provided more realistic feedback, better depth perception and they valued using the real operating equipment, which accompanied the video box trainers.

Laparoscopic virtual simulators are the most well studied virtual surgical simulators. They have demonstrated that novice trainees can acquire psychomotor skills and that these skills can be transferred into real procedures. These simulators have also demonstrated construct, face and concurrent validity, however general surgery residents still feel that the gold standard laparoscopic simulators are more valuable in teaching skills than the newer virtual reality simulators.

Orthopaedic surgery virtual surgical simulators

In 1996 the American Academy of Orthopaedic Surgeons suggested that virtual reality training should be developed for the training of procedural skills in orthopaedics [16]. A Task Force on Virtual Reality was created in 1998 and since then a focus on arthroscopic simulators has been developed [6]. A review of literature has yielded a small number of orthopaedic surgery based virtual simulators since that time. The incorporation of haptics (force-feedback) devices into orthopaedic simulators provides realistic tactile feedback, which was missing in earlier computer based simulators. Virtual arthroscopy simulators for the knee [6,17,65] and shoulders [1,18-20] have been designed using haptics devices. These simulators have demonstrated both face validity [17,20] and construct validity [1,17-19]. Questionnaires have determined that participants felt these simulators would be effective training tools for surgical residents [17,20]. These studies failed to ask if the current virtual simulators were more valuable to training than the current gold standards for arthroscopy, and none of these orthopaedic simulators have demonstrated any transfer of skills to a real patient. An ongoing study yet to be completed involves a randomized multicenter study in which a virtual arthroscopic simulator is being used by half the orthopaedic surgery residents in a program, and the conventional teaching of arthroscopy is being used by the other half [6]. After their training, they will be evaluated with a diagnostic human arthroscopy procedure by their staff using standardized checklists to attempt to provide construct and face validity for the simulator. This is the first orthopaedic virtual simulator attempting to demonstrate the transfer of skills after a randomization of differing arthroscopic training methods.

The only open procedure virtual simulator previously attempted comprised of surgical tools using haptics and three-dimensional graphics of the open abdominal cavity, for suturing techniques [66]. This was an early model, and no attempts for an open orthopaedic simulator have yet been attempted. This may be due to the difficulty of providing a realistic environment and stereoscopic 3D images. Arthroscopy and laparoscopy are simpler as they use a 2D screen to view the procedure being performed in the operating room as well as the skills laboratory, whereas for open procedures a 3D stereoscopic representation is required. It may be difficult to attain a high degree of fidelity for these open procedures, and therefore, they are still relatively novel ideas.

Sawbones

The current gold standard simulator for surgical fixation of the ulna is a Sawbones simulator, where a foam-modeled ulna is used to simulate open surgical fixation in a laboratory setting. Sawbones have less variability than human cadavers, and therefore provide more standardized models with which to learn and practice surgical fixation [67], and have been used to test methods of surgical fixation for the radius [68]. Sawbones have also demonstrated similar external bending and pullout strength properties for screws to real bone [67,69].

When a new virtual simulator is developed with the intent of recreating a certain construct, such as surgical fixation of the ulna, it should be compared to the current gold standard. Procedural measurement or evaluation tools can be utilized to help with this comparison.

Measurement and Evaluation tools

Multiple methods can be used for the evaluation of a surgical trainees performance, such as the Objective Structured Assessment of Technical Skills (OSATS), which was developed in Toronto. OSATS may also be of value for the validation of a surgical simulator.

Objective Structured Assessment of Technical Skills (OSATS)

Surgical education is multifaceted, consisting of knowledge, comfort and skill development over a number of years. The psychomotor (skills) domain is not assessed as thoroughly as the cognitive domain during surgical training. Most evaluations of technical skills are subjective in nature, consisting of summative evaluations performed at the end of rotations based on recall by the attending physician. This subjective evaluation technique has demonstrated low reliability [70]. Structured formal examinations have been revolutionized with the advent of objective structured clinical examinations (OSCE's), which are currently being used to assess students' and residents' clinical performances. Borrowing on the success of the OSCE's, a group in Toronto developed a new assessment tool to evaluate psychomotor skills of surgical trainees. A task-specific checklist for three general surgery procedures was developed for completion during the procedure, followed by a global assessment form to be completed by an examiner immediately after the procedure [71]. Winckel et al (1994) demonstrated good inter-rater reliability and construct validity with their task-specific checklist and detailed global assessment forms while assessing technical skills of surgical trainees in the operating room. The operating room is thought to be the best place to assess the performance of a surgical procedure, however, it is nearly impossible to standardize such a procedure to the extent required to reliably evaluate a trainee's performance. It is difficult to standardize the patient and their injury/condition, and hard to eliminate preceptor or staff interference. It is also costly to the hospital in terms of operating room time, and it may be of ethical concern to have a patient entirely in the hands of a learner. For these reasons, training and evaluating surgical skills may be more appropriately conducted in a laboratory setting, such as a surgical skills laboratory. Surgical skills laboratories are now more commonly used to aid in teaching surgical skill, which most often focus on technical skills in isolation [72,73]. The skills laboratory will allow residents to repeatedly practice skills and allows immediate feedback if educators are present.

A new evaluation tool, based on the checklist and global assessment forms created by Winckle (1994), was developed for use in a surgical skills laboratory setting for evaluating surgical skills. The Objective Structured Assessment of Technical Skills (OSATS) is a tool that is used to assess surgical competency of surgical trainees, or those training to be future consultants [4]. The idea of the OSATS is that instead of performing a recall based subjective evaluation, an evaluator can observe specific domains and give an immediate objective assessment of complex psychomotor skills.

The OSATS as described have two parts. The first is a procedure specific checklist. The checklist identifies the separate actions that expert surgeons have deemed necessary to effectively perform a specific operative procedure. Each step combines motor skills with the cognitive process to produce an action, which is judged to have either been performed correctly (yes), or not performed properly/not performed at all (no). The second part involves the Global Rating Scale (GRS), which evaluates operative performance and competencies. It consists of 7 or 8 domains that are procedure independent. Evaluators complete the assessment upon conclusion of the procedure, rating the level of each competency using a five-point Likert scale with anchored behavioural descriptors. Together, these measurement tools can evaluate from the perception stage to the complex overt response stage, and may be useful up to the adaptation stage of Simpson's taxonomy for psychomotor skills. This format of evaluating technical skills can be used to aid in the evaluation of residents in a skills laboratory setting, and may be used to evaluate new surgical simulators.

When developing any test or assessment tool, it must meet two criteria: it must be valid and reliable. The validity of a test is the concept of whether or not a test measures what it is supposed to measure. Multiple domains of validity were discussed previously. The reliability of a test can best be thought of as the precision of a test. Examples of specific reliability types measured for surgical education measurement and evaluatory tools are inter-rater reliability and internal consistency. Since being introduced, the OSATS have gone through extensive validity and reliability testing [4,74-80].

OSATS Validity

The first trial used to assess the technical skills of surgical trainees using the checklist and GRS in a laboratory setting was in Toronto [4]. Using a within-subjects design study, twenty residents performed the same six procedures on both live and bench models. The live and bench models held a high correlation, and the group stated that a bench model was suitable to evaluate resident surgical skills, avoiding increased costs and ethical concerns of live models. Using a multivariate analysis of variance (MANOVA), the GRS was able to distinguish between residency levels ($p < 0.05$), demonstrating construct validity. Following this study, the same group aimed to provide further evidence of construct validity [77,78]. They demonstrated a significant effect of both the checklist and GRS being able to distinguish between experience levels of the surgical trainees.

The question was then asked if these same stations could be performed elsewhere and maintain their validity. General surgery residents from nine different residency programs in Los Angeles and Chicago participated in this study, once again demonstrating the ability to distinguish between experience levels when using OSATS to assess technical skill bench stations [81]. Other specialties have also demonstrated construct validity using

OSATS, including obstetrics and gynecology [75,76] and urology [80]. In the U.K., a motion analysis device - the Imperial College Surgical Assessment Device (ICSAD) has been validated as a tool to measure surgical performance [82]. The ICSAD was compared to both time to completion and the OSATS for assessing surgical skills [74]. All three of ICSAD, time and GRS, were able to discriminate the experience levels between junior trainees, senior trainees and consultants. The checklist, however, was not able to accurately demonstrate construct validity in this study.

This study also demonstrated concurrent validity, as there was a statistically significant correlation between the ICSAD scores and the GRS ($p < 0.05$) [74]. Regehr (1998) attempted criterion validity by grouping a low number of participants into junior or senior residents [77]. Nineteen faculty members ranked the residents they knew by order of skill level. These resident rankings were then compared to their OSATS scores. Correlation was high for the senior residents, but low for the junior residents. The low correlation with the junior residents was thought to be staff-related, as they did not know the junior residents abilities as well as they knew the senior residents. Face validity, through using a short questionnaire, was also demonstrated in a recent study about OSATS [83]. Through reviewing all these studies, the OSATS demonstrate face, criterion and construct validity.

OSATS Reliability

The reliability of these evaluation tools can best be determined by evaluating their inter-rater reliability using Cohen's kappa or intraclass correlation, and their internal consistency using Cronbach's α. A tool should obtain a significant Cohen's kappa or intraclass correlation, with 1.0 being perfect agreement, to indicate good inter-rater reliability. The internal consistency of an evaluation tool should aim to have a Cronbach's α

above 0.7 [84] resulting in an acceptable scale, or a result of 0.8 for a good scale, indicating its usefulness for accrediting exams.

The reliability of OSATS has been demonstrated through many studies. Martin (1997) using a within-subjects designed study for live and bench models demonstrated an inter-rater reliability of 0.64-0.72, as well as an internal consistency reliability with a Cronbach's α = 0.61-0.74 for the checklists and GRS [4]. Internal consistency using OSATS was also demonstrated to be high in other studies for the checklist and GRS (0.78-0.89) [77,78]. The internal consistency and inter-rater reliability of OSATS was found to be even higher when studied by an obstetrics and gynecology group with an inter-rater reliability of 0.87 for GRS, 0.78-0.98 for the checklists, an α = 0.89 for GRS, and an α = 0.89-0.95 for checklists [76,85]. The checklists are slightly different between the two research groups, with that of the latter group consisting of 5-7 tasks that are evaluated on a five-point Likert scale of how well the task was completed, instead of a simple yes/no checklist. Both groups have suggested that as a result of OSATS' high reliability, only 1 examiner per station is needed for these stations [4,75].

The OSATS have demonstrated face, construct and criterion validity, as well as inter-rater reliability and internal consistency through multiple studies. OSATS have also helped in evaluating the effectiveness of surgical skills curricula in Toronto [72] and Seattle [73], and were used to test new curricula such as a hysteroscopy course [80]. They have also been used to help determine if a cognitive skills curriculum improves learning surgical skills during a course [86]. Surgical trainees and their examiners who have participated in the OSATS format have filled out questionnaires to determine specifically the face validity of OSATS, value of OSATS for training, including the OSATS in Annual training and practicability of OSATS. In all fields, at least 75% of both trainees and examiners answered positively to these domains [83].

The major concern in the literature at this time is that the checklist may be inferior to the global rating scale, for in some cases it fails to discriminate between experience levels [4,74]. This decreased ability to discriminate between groups may be due to the fact that a checklist may have a ceiling effect for certain skills, especially if the stations are meant for a lower level (junior trainee level). It is also difficult for a checklist to portray how well a trainee performed, whereas the GRS can actual say how well a student performed the entire technique, and not just if the trainee knew the steps. It has been suggested that the GRS is more useful for evaluating senior residents if the examiner is an expert and has some training with the evaluation form [77].

Questionnaires

Whether assessing the fidelity of simulators, or determining a participants perceived knowledge, skill and comfort with a procedure, questionnaires are extremely useful tools in research involving virtual surgical simulators.

After the development of a virtual simulator, it is important to assess how well the users believe the surgical simulator recreates the tools, environment and feel of the real procedure. A questionnaire is one of the most valuable tools to determine how well a user feels the simulator has been made and helps determine how valuable it may be in the future. These questionnaires can range from 3-4 Likert style questions to a lengthy, more thorough questionnaire, depending on exactly what the study is attempting to assess. Most studies examining virtual simulators include short questionnaires for the users to complete after they have used the simulator, using either Likert scales or yes/no answers for statements to determine the face validity and effectiveness of a simulator [17,20,57,87-92].

Longer questionnaires have the ability to be more specific and can help determine user-friendliness, training capacity for the simulator and first impressions of the design and their experience with the simulator [30,93]. These questionnaires can also be used to aid with further development of the simulator [94]. Possibly of most importance, these questionnaires may also aid with assessing the fidelity of the simulators. Although no pure fidelity questionnaires were found in recent literature, presence questionnaires have been created. Presence is similar to psychological fidelity and may be defined for surgical simulators as moments during scenarios where the trainee actually feels as though they are in the operating theater [37,87]. Perhaps the most well known are the Steed-Usoh-Slater and Witmer & Singer presence questionnaires [95-97]. These have been used to help determine a users presence in virtual environments, but not in the surgical field.

In order to assess how confident a surgical trainee is with certain procedures, a questionnaire can also be used. When assessing a surgical trainees confidence for a procedure, a questionnaire can be broken down into different categories evaluating the three domains of Bloom’s educational taxonomy: cognitive (knowledge), affective (comfort) and psychomotor (skill). A short questionnaire is sufficient to assess each of these domains, and compare them prior to and after the procedure. Taekman (2004) demonstrated that the use of a simulator significantly improved the user’s confidence overall and within each separate domain [98]. These questionnaires may be useful for evaluating participants on their perceived knowledge, skill and comfort with certain surgical procedures.

The modification and pooling of previously used questionnaires can be used to help evaluate both the simulator itself and surgical trainees when used appropriately. These previous questionnaires can be modified to help one determine the overall fidelity of a simulator, how well it can be used for surgical education, and to determine what further modifications may be required before the simulator can be used as a practice or assessment tool. A questionnaire can also be used in assessing trainees perceived knowledge, skill and comfort with certain procedures to determine what level the trainee feels they have currently attained, and whether practice with the simulator can increase any of these domains.

CHAPTER 3 – METHODS

This chapter contains a description of the simulators, of how the experiment was performed, and how the data was analyzed. This chapter will be divided into eight sections: Description of simulators; Study design; Procedure; Sample description; Description of instruments used for data collection; Data collection; Statistical analyses; and Ethical considerations.

Description of the Simulators

Sawbones Simulator

Sawbones is a brand name of synthetic modeled bones, specifically designed for use in motor skills exercises. They are used internationally at procedural fracture fixation courses for both learning and practicing psychomotor skills. Sawbones are used instead of cadavers due to availability, cost, ethical concerns, standardization, ease of clean up and lack of contact with potentially biohazardous material. Sawbones of a hand and forearm are seen in Illustration 1.

Illustration 1. Forearm Sawbones (ulna bottom, radius top)

The Sawbones and required resources were donated by Synthes (Canada) Ltd for the Sawbones simulator procedure of surgical fixation of the ulna (model 1017, Sawbones; Pacific Research Laboratories, Vashon, Washington). These are made of rigid foam and are most commonly used for external fixation procedures, limited total joint replacements, and internal fixation (as for this study) [99]. The ulna was used alone and was placed in a vice, as seen in Illustration 2. A line was drawn on the ulna to simulate a fracture line, and the procedure was then carried out. Sawbones simulators are the gold standard for learning and practicing fracture fixation of the ulna.

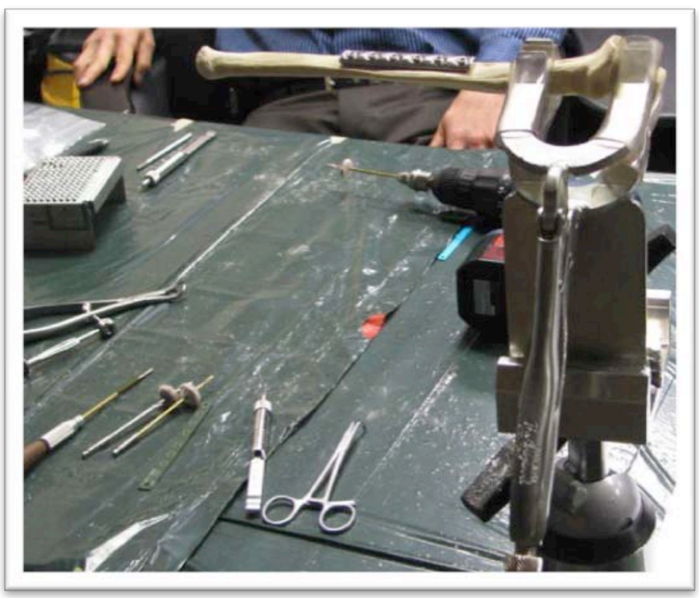


Illustration 2. Surgical fixation of the ulna in vice

Virtual Bone Simulator

A virtual simulator for fracture fixation of the ulna was developed in collaboration with the University of Calgary Department of Electrical and Computer Engineering. From May 2008 to April 2009, both an undergraduate and a graduate engineering student, neither with prior experience in virtual reality modeling, created the simulator. Their supervisor and a PhD student with prior experience in this field were available to aid these students when needed. This simulator consists of the distal arm, more specifically the fractured bone (ulna), with an option to allow skin to be seen, or turned transparent. Tabs displayed around the top and sides of the screen provided a series of optional tools to use during surgical fixation of the ulna (Illustration 3). A haptics device (PHANTOM 1.5/6DOF, SensAble Technologies Inc. MA, USA) that provides realistic force-feedback during the procedure was used to allow the user to move the tools around the screen and feel resistance when the bone was touched (Illustration 3).

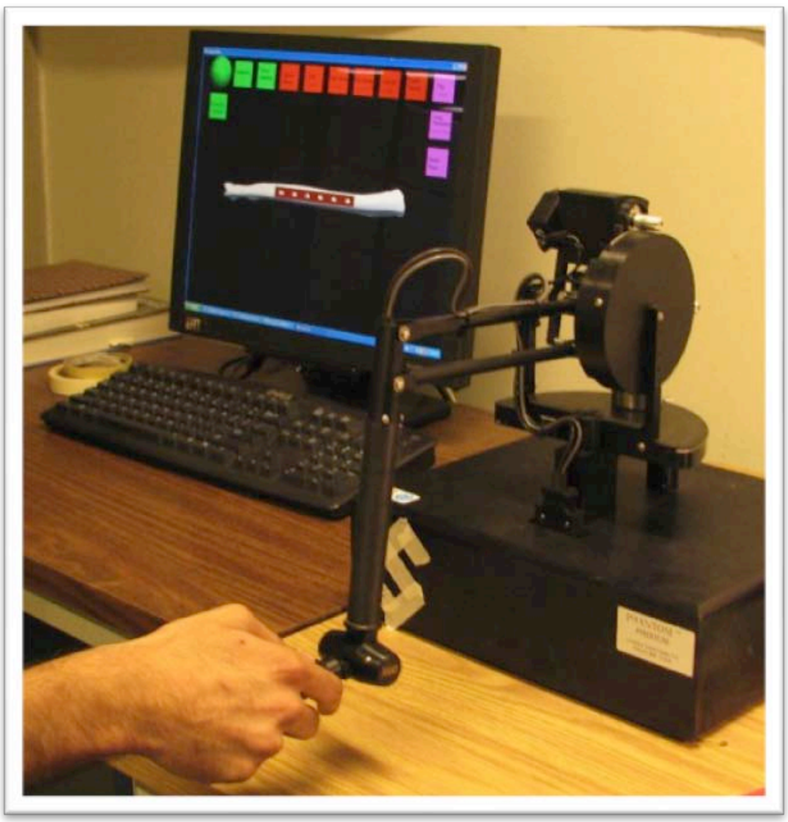


Illustration 3. Virtual Simulator: Display (left) and Haptics device (right)

The orthopaedic surgery resident in charge of this study (author) was available at all times to answer questions about any task of the procedure or tools involved during the simulators development. This was to ensure that the virtual simulator reflected the real surgical fixation of an ulna procedure as much as possible. The Hardware used was a Pentium 4 CPU, 3.00 GHZ, 2GB of RAM. The operating system was Windows XP. Computer Language was C++ and the compiler was Visual Studio 2005. The software used to create this simulator were OpenGL (Open Graphics Library) and Open Haptics API (Application Programming Interface).

Study Design

A stratified randomized within-subjects design was used to compare surgical fixation of the ulna between the virtual simulator and the Sawbones simulator. Participants were stratified according to experience level and gender. The participants were randomized by a computer generated number to either begin with the Sawbones simulator or virtual simulator, and then completed a surgical fixation of the ulna using that specific simulator. After the first simulator procedure, they then completed the same procedure using the other simulator. A post-questionnaire was completed after each procedure to assess the fidelity of each specific simulator.

Sample Description

Sample size

In a review of literature for the use of virtual simulation in resident education, the number of participants used varied from 16-43. Only one previously found study provided a sample size calculation [100]. Using the same values of an alpha of 0.05, beta value of 0.20, determining a 20% difference between groups (based on checklist scores of 90% and 70%), and a SD of 15%, a sample size of 9 participants per group was determined [101]. Twenty-four orthopaedic surgery residents were available and volunteered for the study. Two participated in the pilot project, and the other 22 participated in the full study, 11 in each group.

Participants

The pilot study volunteers consisted of two first year medical students, two third year orthopaedic residents and one staff orthopaedic surgeon, all from the University of Calgary (U of C) and Alberta Health Services (AHS). All participants were contacted via email regarding this study, and volunteered once the study process was disclosed. The medical students and residents were all between the ages of 25-29, while the staff member was over 40 years of age. Among the participants were two females and three males. All participants signed an informed consent prior to the procedures (Appendix 1). Unique to the pilot study, both participants and examiners were asked to evaluate the questionnaires and checklists after they completed the procedures, and to identify any concerns with the methods of the study.

The full study participants were all residents from the orthopaedic surgery training program at the University of Calgary. All available orthopaedic surgery residents at the U of C volunteered for this study. Calgary has 26 orthopaedic surgery residents, and only four residents were excluded from this study for the following reasons: two had participated in the pilot project; one was not in Canada at the time of the study; and the last resident was the author of this thesis. Of these 22 participants, all participants signed an informed consent prior to beginning the procedures. There were four post-graduate year (PGY) -5's, three PGY-4's, three PGY-3's, seven PGY-2's and five PGY-1's. Eight participants were female, two were left-handed and their ages varied from 25-40 years old.

Examiners

Five separate examiners volunteered their time for this study. Examiners had all successfully completed an AO (Arbeitsgemeinschaft für Osteosynthesefragen) fracture fixation skills course for standardization purposes, where they learned and practiced proper techniques for surgical fixations. The pilot project had two examiners, both were physicians who had prior experience evaluating simulators with checklists and global rating scales and both work for Alberta Health Services (AHS). One examiner was used for each simulator. The full study used both of the previous examiners (from pilot study), another orthopaedic surgeon from the AHS and two orthopaedic surgery trained fellows. Two examiners evaluated ten participants on the virtual simulator and six participants on the sawbones simulator. Due to the unavailability of more examiners, all other participants had only one examiner per station.

Examiner Training

Checklists and the global rating scales (GRS) were provided to the examiners 5 days prior to the study date and they were allowed to contact the researcher with any questions or concerns. Examiners were also trained on the examination day for 10 minutes. This consisted of reviewing both evaluation tools, and answering any questions or addressing any concerns the examiners had about the process or their role as an examiner.

Procedure

The procedures were performed at the U of C Department of Electrical and Computer Engineering in two separate rooms, one for each simulator. The first room contained the newly created virtual simulator with either one or two evaluators. Instructions for the procedure were written on the bench, which included specific directions about how to select and use tools on the virtual simulator. The second room contained the Sawbones simulator with all required instruments and one or two evaluators. Once again, instructions for the procedure were written and placed on the procedure table.

Pilot study

A pilot study was performed to assess the simulators, as well as to evaluate the measurement tools: checklist, GRS and questionnaires. The participants consisted of two medical students, two PGY-3 residents and one staff orthopaedic surgeon. Participants first went through a short orientation to the experiment and then completed both procedures. Once the pilot study was evaluated, minor changes were made to the checklist, global rating scale and the pre-procedure questionnaires.

Full Study

For the full study, participants were given an orientation to how this experiment would proceed at two separate times. The first was at an introduction seminar during a mandatory academic half-day for orthopaedic residents, three weeks prior to the experiment. The second orientation was performed when the participants arrived on the study date. They were given a quick introduction to the simulators and the study. Questions were answered, participants signed a consent form and then completed a pre-procedure questionnaire. Participants were stratified for experience level and sex, and then randomized to begin with either the Sawbones simulator procedure or the virtual simulator procedure by using a computer generated randomized number [102]. Group 1 (called the Sawbones used first group) performed the fracture fixation of the ulna with the Sawbones simulator first, using the same tools normally found in the operating room and at procedural skills training courses. This involved using a power drill, screwdriver, tap, drill guides, depth gauge, screws and a surgical plate (Illustration 4), which along with the ulna Sawbones were donated by Synthes (Canada) Ltd.

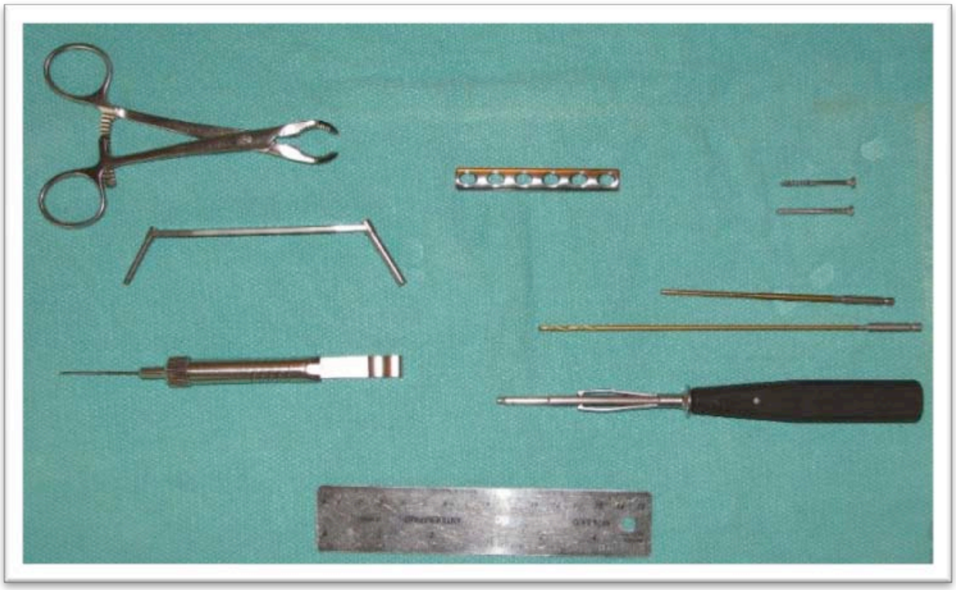


Illustration 4. Operating tools used for international Sawbones courses

The participants were given a post-procedure simulator specific questionnaire after completion of the procedure. After completing the questionnaire, participants were then asked to perform the same procedure using the virtual simulator with a haptic device (providing force feedback). This consisted of a computer screen and a handheld device that provides the user with simulated haptics that would be expected from completing an ulna fixation on a real patient (Illustration 3). A post-procedure simulator specific questionnaire was given to the participant after completion of the procedure. Group 2 (called the virtual used first group) first performed the procedure using the virtual simulator with haptics device, followed by the Sawbones simulator.

This virtual reality simulator was a new form of simulation to all the participants, and the majority of them had already used some form of Sawbones simulation before. Participants were therefore given ten minutes to learn the basic controls of the virtual simulator before performing the procedure. They had a standardized three-minute walkthrough of how to use all tools on the simulator, and then had the remaining time to become comfortable using these tools. If any residents were uncomfortable with the Sawbones simulator and tools, they were also given ten minutes to become comfortable with the instruments and the feel of the Sawbones. Examiners were allowed to instruct participants on how to use the tools during the practice session for both simulators, but were not allowed to instruct in what order specific tasks were to be completed. After the practice time, examiners were no longer allowed to discuss how to use the tools, except in the case of the virtual procedure where certain key combinations were sometimes confusing.

During the procedures, two independent examiners marked the participants on their performance using a modified itemized checklist (including number of errors and time to completion) and a global rating scale for each of the procedures.

Study date

The pilot study was performed on March 26th, 2009. The set up occurred three days prior for the virtual simulator. The virtual simulator had to be moved from one location to another, so once it was set up in the examination room, the simulator was thoroughly re-tested to ensure it was working well. The simulator was set up so that no parts of the computer (except the display), keyboard (except spacebar), or tables were seen – all covered with black cardboard or black plastic bag. The day of the pilot study, organization began for the Sawbones room an hour prior to the participants arriving, to ensure the set up for this procedure was complete.

The full study was performed on April 2nd, 2009. The simulator had not been moved, and the software was not changed from the time of the pilot study. Once again, three days prior to the study, the same set up was completed for the virtual simulator. The day of the study, the Sawbones room was set up as described above, with all Sawbones and equipment in the room.

The set up of the virtual simulator in between participants was the responsibility of the examiner in that room. The simulator had to be reset each time, and then tested to ensure the haptics was functioning appropriately for each participant. The set up of the Sawbones simulator was the responsibility of the examiner in that room as well, who would set up a new ulna Sawbone for each participant and ensured all required tools were available.

Description of instruments

Questionnaires were developed to both acquire participants' perceived confidence in surgical fixation of an ulna and to assess the fidelity of both simulators. The evaluation tools used to assess the surgical skills of the participants are the checklists and global rating scales. The examiners also scored total errors per procedure and total time to completion of the procedure. Their descriptions follow.

Questionnaires

A pre-procedure questionnaire was developed to collect general information about the participants and their familiarity with the surgical fixation procedure. The first ten questions pertained to demographics and their previous experience with simulators and video games (Appendix 2). On this same questionnaire, ten questions regarding the participants' own perceived knowledge, skill and comfort with surgical fixation of an ulna were assessed using a 5-point Likert scale. A second set of the exact same ten questions also appeared on the virtual simulator post-procedure questionnaire, to determine if the participants' perception of their self-assessed skill level changed after using the simulators to complete the surgical fixation of the ulna.

Two simulator specific questionnaires were developed and given to the participants once surgical fixation of the ulna had been completed with that specific simulator. These were used to assess the residents' perception of certain domains of fidelity for both the Sawbones simulator (Appendix 3) and virtual reality simulators separately (Appendix 4). Each questionnaire contained the same questions, but was specific to the simulator used. The questions were designed to assess the physical and psychological fidelity of both the virtual and Sawbones simulators.

Physical fidelity can be defined as the extent to which the physical simulator looks, sounds and feels like that of the operational environment in terms of visual displays and controls [36]. The virtual model has no audio component and consequently the audio portion of the fidelity was unable to be assessed. The physical fidelity aspects were divided into two domains on the questionnaire: 1) environment: the extent to which the simulator duplicates motion, visual and other sensory information from the true environment; and 2) equipment: the degree to which the simulator duplicates the appearance and feel of the real system.

The psychological fidelity had one measured domain: the degree to which simulation replicates psychological factors, such as stress and fear, which can be experienced in the real environment [37].

A literature review of medically related questionnaires for virtual simulator fidelity provided many types of questions. The questions most closely associated with this type of procedure were modified and used, and some new questions relating to our specific objectives were also created.

All three questionnaires were assessed during the pilot study to ensure no grammatical problems existed and that all questions were meaningfully related to the ulna fixation procedure. Orthopaedic surgery and medical education experts, who have worked with these types of questionnaires before, also evaluated the questionnaires to help assess their face validity.

Procedure Skills Checklist and Global Rating Scale

To assess the quality of their performance, data from the participants was acquired using a checklist and global rating scale (GRS) designed specifically for the simulators used in this study. They are, however, based on the checklists and global rating scales used during the Objective Structured Assessment of Technical Skills stations first described in Toronto [4]. The versions of these evaluation tools for surgical fixation of an ulna were modified from the OSATS station for surgical fixation of the radius (forearm bone).

The modified checklist was used to evaluate correct tool usage, correct order of tools used and if tasks were performed accurately. This checklist is composed of 15 specific tasks that need to be correctly completed in order to perform an accurate surgical fixation of the ulna (Appendix 5). Other objective measurements on the checklist were total errors and time to completion of procedure (seconds) as recorded by the examiners.

The modified GRS consists of 6 domains: principles of fracture fixation, definitive fixation, flow of operation, instrument handling, time & motion and overall performance.

The examiners filled out the global rating scale after the procedure was completed, rating the level of each competency on a 5-point Likert scale (Appendix 6).

Data collection

Data Sources

Data was gathered from the participants at three different stages: (1) pre-procedure questionnaires; (2) checklists and global rating scales during the surgical procedures of both simulators; and (3) simulator specific post-procedure questionnaires.

Pre-procedure Questionnaire

Each participant was given a questionnaire to be filled out independently prior to completing the procedures (Appendix 2). The questionnaire is described in the above section. Data was collected from each participant for group demographics and to enable a comparison of knowledge, skill and comfort levels among groups both before and after the procedures. The last ten questions (part B) were assessed using a 5-point Likert scale (1 = “inferior”, 3 = “average”, 5 = “excellent”).

Procedural Checklist and Global rating scale

Each participant completed two procedures: surgical fixation of the ulna using (1) Sawbones simulator and (2) virtual simulator. During theses procedures, the participants were evaluated by either one or two independent examiners who filled out the itemized 15-task specific checklists. Each specific task was evaluated, and if it was performed correctly one full point was given. If the task was performed poorly, wrong or was missed, then zero points were given. No half points were used. During the procedure, the examiners also recorded the total amount of errors occurring during the procedure and the total time taken to complete the procedure. After the procedure was completed, the participants were evaluated by the examiner(s) using a 6 domain GRS, based on a 5-point Likert scale. Anchored behavioural descriptors were present for points 1,3 and 5 (1 was rated as “poor”, 3 was “average” and 5 was “excellent”).

Post-procedure Questionnaires

Immediately after each procedure, the resident was given a simulator specific questionnaire. The base questions were the same for each questionnaire, and designed to determine the fidelity of that specific simulator compared to the participants’ experience with the same procedure in real life. A 5-point Likert scale was used where 1 was rated as “strongly disagree”, 3 was “neutral” and 5 was “strongly agree”.

Statistical Analyses

The statistical analyses and derived findings were obtained using SPSS v.17.0 (SPSS Inc. Chicago, Illinois). The pre-procedure questionnaire, procedural measurement tools (checklist, GRS, total errors, time to completion), post-procedure questionnaires and feasibility (cost analysis and exit interviews) were all evaluated separately. Statistical significance was reached if $p < 0.05$, and effect size differences (Cohen’s d) were calculated when significance was reached. When data did not adhere to the three basic assumptions of: 1) data having a normal distribution; 2) equal variance among groups; and

3) independent observations; then non-parametric testing was conducted using either a Mann-Whitney U tests or a Kruskall Wallis test, as indicated.

Sample Description

The participant's personal data was obtained from the first ten questions of the pre-procedure questionnaire data. These were used to make comparisons between groups (i.e., simulation used first, experience level).

Pre-procedure Questionnaire

The pre-questionnaire also contained 10 self-assessment questions of the participants' knowledge, skill and comfort for surgical fixation of the ulna. The same ten questions were asked again after the procedures, and these were compared to determine if the participants perceived any improvement in knowledge, skill or comfort for surgical fixation of the ulna. Paired samples t-tests were conducted to compare the pre-procedure and post-procedure total scores, as well as scores from the separate domains (knowledge, skill, comfort). Separate paired samples t-tests were also conducted for each level of experience comparing pre- and post-procedure questionnaires. Experience level was divided into separate post-graduate years (PGY's separately) and into junior (PGY1,2), or senior (PGY-3,4,5) residents. Effect size differences were evaluated using Cohen's d: with $d = 0.2$ to 0.49 being a small effect, $d = 0.5$-0.79 being a medium effect, and $d \geq 0.80$ being a large effect size difference [103].

Independent t-tests were conducted to assess any differences among group 1 (Sawbones simulator used first) and group 2 (virtual simulator used first). To determine if this questionnaire could accurately distinguish between experience levels, a one-way analysis of variance (ANOVA) was conducted between the PGY levels. An independent t-test was used to distinguish any difference in means between the experience levels of junior or senior residents. The same tests were also conducted on the same groups to assess the individual domains of knowledge, skill and comfort separately.

Procedural Measurement tools

A descriptive analysis was conducted using the procedural measurement data of the checklist, global rating scale (GRS), total errors and time to completion for each simulator. Mean scores, standard deviations, minimum and maximum scores, skewness and kurtosis were calculated. The data was defined as skewed to a significant degree if the score was at two standard errors of the skewness (SES) or more, and the data was kurtosed if the score was at two standard errors of the kurtosis (SEK) or more [104]. When this occurred, non-parametric tests were used when evaluating the measured tools.

A direct comparison of total measured scores between the Sawbones and virtual simulators was assessed. Four paired samples t-test were used to compare the measured parameters, to determine if there were any significant differences between the virtual simulator and Sawbones simulator scores. Paired t-tests were also conducted to evaluate any significant difference in means between simulators at each separate experience levels (both PGY's separate and junior versus senior).

The difference in means for the procedural measured data was determined by conducting an independent t-test to compare group 1 (Sawbones simulator used first) and group 2 (virtual simulator used first). A Mann-Whitney U test was used when the data was significantly skewed or kurtosed.

Construct validity of the simulators were evaluated by determining if they could differentiate between experience levels. Multiple one-way ANOVA's were conducted to analyze the difference of means between experience levels (PGY's separately) for measured scores. A post-hoc analysis using Tukey's B was performed if significance was found ($p < 0.05$). Independent t-tests were used to compare experience levels of junior and senior residents against the procedural measured data (checklist, GRS, total errors and time to completion). A multivariate analysis of variance (MANOVA) added further information, so was also conducted.

Criterion related validity was determined by the Pearson product moment correlations (r) within and between simulators, which identifies the relationship within and between the evaluation tools. Both simulated procedures evaluated the surgical fixation of an ulna, with the Sawbones procedure being the gold standard. Within simulators analysis looked at the correlation between checklists and GRS of that specific simulator. Between simulators looked at the correlation of checklists and GRS between both simulators.

The internal consistency reliability of the checklist and global rating scales were determined by Cronbach's α. An acceptable tool has a reliability of $\alpha \geq 0.70$ [84], and a good tool has $\alpha \geq 0.80$. Inter-rater agreement between the independent evaluators' checklist and GRS scores were assessed with a Cohen's Kappa coefficient, with complete agreement being $\kappa = 1.0$, no agreement being $\kappa \leq 0$. The total errors were unable to be calculated due to a lack of value points, therefore an intraclass coefficient was used to calculate inter-rater agreement for checklists, GRS and total errors for each simulator. Once again, a value of 1.0 reflects perfect agreement, and 0 reflects no agreement.

Post-procedure Questionnaire

A descriptive analysis was conducted for the post-procedure questionnaire data. Mean scores, standard deviations, minimum and maximum scores, skewness and kurtosis were calculated.

Two separate simulator specific questionnaires were created to determine the fidelity of each simulator. The fidelities of each of the three fidelity domains (environment, equipment and psychological) were determined by assessing the means of each domain using a percentage score. The domains were compared between groups, both between which simulator was used first, and between experience levels. The simulator used first groups were either the Sawbones simulator (group 1) or the virtual simulator (group 2). Experience level was divided into separate post-graduate years (PGY's separately) and into junior (PGY-1,2), or senior (PGY-3,4,5) residents.

Independent t-tests were used to compare fidelities of each of the three fidelity domains for each simulator between the simulator first used groups. Multiple one-way analysis of variance tests (ANOVA's) were used to compare the same fidelity domains between the separate PGY experience levels. To ensure no statistical differences were missed, or incorrectly calculated, a MANOVA was used to compare means of fidelity domains between both simulator used first groups, and experience levels. Independent t-tests were used to compare fidelities of each of the three fidelity domains for each simulator between the junior and senior experience levels.

The fidelities of the two simulators were compared directly using paired samples t-tests. The fidelity domains of each simulator were compared to assess which simulator had a higher mean score, and if those means were significantly different.

Three separate paired samples t-tests were conducted for each simulator (environment versus equipment, environment versus psychological, and equipment versus psychological). These were used to determine if any differences existed between the fidelity domains within each specific simulator.

The internal consistency for each fidelity domain of the questionnaires was also assessed to ensure the same construct was measured for each domain. Each domain was measured separately and all together.

A descriptive table was developed for each separate question from the questionnaire. A paired t-test was conducted to compare the Sawbones and virtual simulators' average of each question.

At the end of each questionnaire, short answer questions regarding the simulators 1) strengths, 2) weaknesses, 3) potential changes and 4) future benefits; were answered by each participant. This information may be useful to further construct a higher fidelity simulator that can be used for surgical fixations. The answers were divided into themes, and placed into a table to determine what participants thought of the simulators.

Feasibility

A cost analysis of both simulators was performed to assess the feasibility of developing a virtual surgical fixation of an ulna. Exit interviews were also conducted with each creator of the virtual simulator to unveil their thoughts about its development. The costs and exit interviews may help assess if it is feasible for another group from another center to be able to create a high fidelity simulator of their own.

Ethical Considerations

All participants who were approached to participate in this study were provided with informed consent in order to participate. Ethical approval was obtained from the University of Calgary Health Research Ethics Board prior to commencing with the procedures. Participants were assigned identification (ID) numbers when they consented. These ID numbers appeared on their questionnaires and evaluation forms to ensure confidentiality. The questionnaires and the data collected from the observers and the virtual models were all placed under a password-protected computer or locked cabinet. All data is available only to the researchers involved directly with this project and will be disposed of five years after completion of the study.

CHAPTER 4 - RESULTS

This chapter provides descriptive data of the participants and the analysis of each measurement tool: 1) pre-procedure questionnaire, 2) procedural measurement tools (checklist, global rating scale, total error and time to completion); and 3) post-procedure questionnaires. Lastly, it will provide a cost analysis of developing a virtual simulator for surgical fixation of the ulna, along with exit interviews from the developers to assess the feasibility of creating a virtual ulna fracture fixation simulator.

Sample Description

Twenty-two orthopaedic surgery residents participated in the full study, from all five years of the orthopaedic surgery residency program offered at the University of Calgary. Post-graduate year (PGY), sex, handedness and age are reported in Table 1, comparing the randomized groups of Sawbones simulator used first (group 1) and virtual simulator used first (group 2). The only noted difference was that no PGY-4's were in group 1 and no PGY-3's were in group 2. Otherwise the groups were relatively evenly distributed by sex, handedness and age categories.

Table 1. Description of Groups

	Group 1 N	Group 1 (%)	Group 2 N	Group 2 (%)
PGY-5	2	(18)	2	(18)
PGY-4	0	(0)	3	(27)
PGY-3	3	(27)	0	(0)
PGY-2	4	(36)	3	(27)
PGY-1	2	(18)	3	(27)
Total:	11	(100)	11	(100)
# Female	3	(27)	5	(45)
Left-Handed	1	(9)	1	(9)
Ages (#/group)				
25-29	7	(64)	6	(55)
30-34	2	(18)	4	(36)
35-40	2	(18)	1	(9)

These participants were grouped based on whether they were junior (PGY-1,2) or senior (PGY-3,4,5) residents (Table 2). There were two extra participants in the junior resident group while the senior resident group had participants that were in older age categories.

Table 2. Description of Groups (Junior versus Senior residents)

	Junior N	Junior (%)	Senior N	Senior (%)
Number each group	12	(100)	10	(100)
Virtual model used first	6	(50)	5	(50)
# Female	5	(42)	3	(30)
Left Handed	0	(0)	2	(20)
Ages (#/group)				
25-29	9	(75)	4	(40)
30-34	3	(25)	3	(30)
35-40	0	(0)	3	(30)

The questionnaire also asked the participants about their video game experience. All residents had previous experience with video games, and a trend of more junior residents still playing video games was noted versus senior residents. Video game experience, however, had no influence over performance on any procedural measurement or their opinions of simulator fidelity.

Pre-procedure Questionnaire

The participants completed a pre-procedure questionnaire, which had ten questions assessing their perceived knowledge, skill and comfort regarding surgical fixation of the ulna. After they had completed the procedures, they once again filled out the same set of questions. This was used to assess if any self-assessed improvement was perceived after completion of a practice and testing session for each surgical fixation of the ulna simulator. A paired samples t-test compared the total means from the pre-procedure questionnaire to the post-procedure questionnaire. The post-procedure questionnaire (M = 75.6, SD = 18.36) had a significantly higher mean than the pre-questionnaire (M = 71.6, SD = 22.13) $t(1,22) = -2.395$, $p < 0.05$, $d = 0.20$. When looking at each individual PGY's results from the pre-procedure and post-procedure scores, the PGY-1's had the only significant difference in means (Table 3).

Table 3. Differences in means by experience level between pre and post procedure questionnaires

	Pre-procedure mean % (SD)	Post-procedure mean % (SD)	*p* value	Effect size (*d*)
PGY-1	41.6 (10.90)	57.2 (14.18)	**0.031**	**1.3**
PGY-2	66.6 (9.43)	66.9 (9.99)	0.356	0.03
PGY-3	74.0 (6.93)	76.7 (4.61)	0.184	0.53
PGY-4	92.7 (7.02)	93.3 (6.11)	0.423	0.10
PGY-5	100.00	100.00		0

The pre and post-procedure questionnaire means were also compared among junior and senior residents. No statistical difference was found between means for the senior residents. The junior resident post-procedure questionnaire (*M* = 62.8, SD = 12.34) yielded higher means than the pre-questionnaire (*M* = 56.2, SD = 16.03), $t(1,12) = -2.262$, $p < 0.05$, $d = 0.48$.

Knowledge, skill and comfort domains were also analyzed separately. A paired samples t-test was conducted to compare pre-procedure scores to post-procedure scores for each separate domain. A significant difference was noted between pre-procedure (*M* = 69.7, SD = 23.34) and post-procedure (*M* = 73.9, SD = 19.40) for the skill domain, $t(1,22) = 2.134$, $p < 0.05$, $d = 0.18$. A significant difference was also noted between pre-procedure (*M* = 71.4, SD = 23.00) and post-procedure (*M* = 75.9, SD = 18.75) for the comfort domain, $t(1,22) = -2.339$, $p < 0.05$, $d = 0.22$.

The pre-procedure and post-procedure scores on the knowledge, skill and comfort domains were also compared to each other using paired t-tests for each separate experience level (Tables 4-6). Two different groups of experience levels were assessed. The first group was all post-graduate years (PGY 1-5) compared separately, and the second group was the junior and senior level residents.

No significant difference in the knowledge domain existed between pre and post questionnaire scores at any experience level. The PGY-1's reported their skill pre-procedure (*M* = 37.3, SD = 10.11) as a significantly lower score than their skill post-procedure (*M* = 54.7, SD = 19.67), $t(1,5) = -3.20$, $p < 0.05$, $d = 1.16$ (Table 5). The PGY-1's also perceived their comfort pre-procedure (*M* = 40.0, SD = 9.35) as a significantly lower score than their comfort post-procedure (*M* = 57.0 SD = 13.51), $t(1,5) = -3.157$, $p <$

0.05, d = 1.48 (Table 6). No other significant differences were found among any experience level for the domains.

Table 4. Differences in means by experience level between pre/post procedure questionnaires assessing Knowledge

	Pre-procedure mean % (SD)	Post-procedure mean % (SD)	*p* value	Effect size (*d*)
PGY-1	48.0 (15.92)	60.0 (10.54)	0.152	0.89
PGY-2	66.7 (9.43)	68.6 (13.17)	0.356	0.17
PGY-3	75.6 (7.70)	75.6 (3.85)	0.999	0.01
PGY-4	95.6 (7.70)	95.6 (7.70)		0
PGY-5	100.00	100.00		0

Table 5. Differences in means by experience level between pre/post procedure questionnaires assessing Abilities

	Pre-procedure mean % (SD)	Post-procedure mean % (SD)	*p* value	Effect size (*d*)
PGY-1	37.3 (10.11)	54.7 (19.67)	**0.033**	**1.16**
PGY-2	65.7 (8.10)	65.7 (8.10)		0
PGY-3	71.1 (7.70)	75.6 (3.85)	0.184	0.74
PGY-4	91.1 (7.70)	88.9 (10.18)	0.423	0.26
PGY-5	100.00	100.00		0

Table 6. Differences in means by experience level between pre/post procedure questionnaires assessing Comfort

	Pre-procedure mean % (SD)	Post-procedure mean % (SD)	*p* value	Effect size (*d*)
PGY-1	40.0 (9.35)	57.0 (13.51)	**0.034**	**1.48**
PGY-2	67.1 (11.85)	66.4 (10.70)	0.356	0.06
PGY-3	75.0 (10.00)	78.3 (5.77)	0.423	0.47
PGY-4	91.7 (7.64)	95.0 (5.00)	0.184	0.51
PGY-5	100.00	100.00		0

The pre-procedure and post-procedure scores of the knowledge, skill and comfort domains were then compared to each other using paired t-tests for junior and senior residents separately (Table 7). No significant differences were noted between domains for either junior or senior residents.

Table 7. Differences in means by experience level between pre/post procedure questionnaires assessing knowledge, skill and comfort

	Pre-procedure mean % (SD)	Post-procedure mean % (SD)	*p* value	Effect size (*d*)
Junior knowledge	58.9 (15.26)	65.0 (12.43)	0.085	0.43
Senior knowledge	91.3 (12.19)	91.3 (11.78)		0
Junior skill	53.9 (16.93)	61.1 (14.45)	0.053	0.47
Senior skill	88.7 (13.71)	89.3 (11.84)	0.591	0.06
Junior comfort	55.8 (17.41)	62.5 (12.34)	0.075	0.44
Senior comfort	90.0 (12.47)	92.0 (10.33)	0.104	0.18

Independent t-tests were conducted to compare pre and post-procedure questionnaires scores between group 1 and 2 (simulator used first groups). There were no significant differences noted between the mean scores of any domains.

Construct validity

To assess if this questionnaire could accurately distinguish between experience levels, an analysis of variance (ANOVA) was conducted. A significant difference was found between experience levels for both the pre-procedure questionnaire: $F\ (4,22) = 32.67$, $p < 0.001$; and post-procedure questionnaire: $F\ (4,22) = 15.25$, $p < 0.001$. A Tukey B post hoc analysis identified three distinct groups for the pre-questionnaire data: PGY-1 < PGY-2 & 3 < PGY-4 & 5 (Figure 1). For the post-procedure questionnaire data, four groups were identified: PGY-1 & 2 < PGY-2 & 3 < PGY-3 & 4 < PGY-4 & 5 (Table 8), although subjectively it may be divided into two groups: PGY-1, 2, 3 < PGY-4 & 5 (Figure 2).

Figure 1. Pre-questionnaire means for each experience level

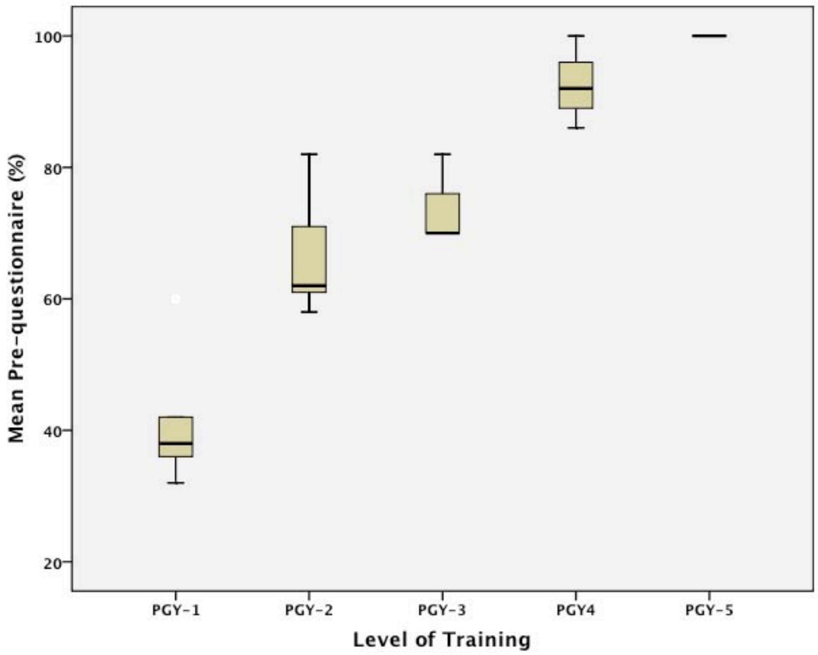
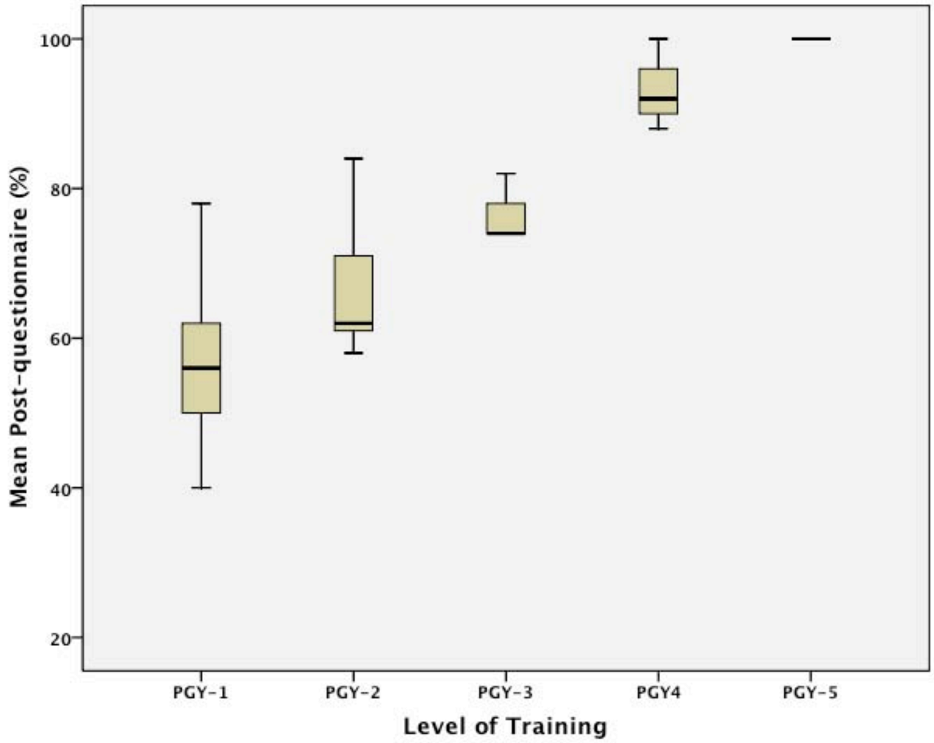


Table 8. Post hoc (Tukey's B) analysis of post-procedure questionnaire between experience levels (subset alpha = 0.05)

Level of training	N	Group 1 Mean	Group 2 Mean	Group 3 Mean	Group 4 Mean
PGY-1	5	57.2			
PGY-2	7	66.9	66.9		
PGY-3	3		76.7	76.7	
PGY-4	3			93.3	93.3
PGY-5	4				100.0

Figure 2. Post-questionnaire means for each experience level



An independent t-test was conducted to assess if any significant difference in scores existed between junior and senior residents for the pre and post-procedure questionnaires. The junior residents had a significantly lower score (*M* = 56.2, SD = 16.03) than the senior residents (*M* = 90.0, SD = 12.40) on the pre-procedure questionnaire, t(1,22) = -5.45 $p <$ 0.001, d = 2.41. The junior residents also had a significantly lower mean score (*M* = 62.8, SD = 12.34) than the senior residents (*M* = 91.0, SD = 10.92) on the post-procedure questionnaire, t(1,22) = -5.61, $p < 0.001$, $d = 2.56$.

An ANOVA was conducted to assess construct validity of the separate domains of knowledge, skill and comfort among post-graduate year's (Table 9). Each domain demonstrated a significant difference between experience levels.

Table 9. ANOVA to assess construct validity of simulator from pre-questionnaire domains (pre = pre-procedure questionnaire, post = post-procedure questionnaire)

	Sum of Squares	Df	Mean Square	F	*P* value
Knowledge pre					
Between groups	7858.72	4	1964.68	18.725	**<0.001**
Within Groups	1783.70	17	104.92		
Total	9642.42	21			
Skill pre					
Between groups	10402.85	4	2600.71	42.529	**<0.001**
Within Groups	1039.58	17	61.15		
Total	11442.42	21			
Comfort pre					
Between groups	9599.57	4	2399.89	27.027	**<0.001**
Within Groups	1509.52	17	88.80		
Total	11109.09	21			
Knowledge post					
Between groups	5097.45	4	1274.36	13.259	**<0.001**
Within Groups	1633.86	17	96.11		
Total	6731.31	21			
Skill post					
Between groups	5725.68	4	1431.42	11.176	**<0.001**
Within Groups	2177.35	17	128.08		
Total	7903.03	21			
Comfort post					
Between groups	5849.44	4	1462.36	16.223	**<0.001**
Within Groups	1532.38	17	90.14		
Total	7381.82	21			

For each domain, a Tukey's B post-hoc analysis was conducted for the pre-questionnaire scores. The knowledge domain for the pre-questionnaire identified 3 group mean subsets: PGY-1,2 < PGY-2,3 < PGY-4,5 (Table 10). The skill domain for the pre-questionnaire identified 3 distinct group mean subsets: PGY-1 < PGY-2,3 < PGY-4,5 (Table 11). The analysis of the comfort domain for the pre-questionnaire resulted in 4 subset groups: PGY-1 < PGY-2,3 < PGY-3,4 < PGY-4,5 (Table 12).

Table 10. Post hoc (Tukey's B) analysis for Knowledge domain of pre-questionnaire between experience levels (subset alpha = 0.05)

Level of training	N	Group 1 Mean	Group 2 Mean	Group 3 Mean
PGY-1	5	48.0		
PGY-2	7	66.7	66.7	
PGY-3	3		75.6	
PGY-4	3			95.6
PGY-5	4			100.0

Table 11. Post hoc (Tukey's B) analysis for Skill domain of pre-procedure questionnaire between experience levels (subset alpha = 0.05)

Level of training	N	Group 1 Mean	Group 2 Mean	Group 3 Mean
PGY-1	5	37.3		
PGY-2	7		65.7	
PGY-3	3		71.1	
PGY-4	3			91.1
PGY-5	4			100.0

Table 12. Post hoc (Tukey's B) analysis for Comfort domain of pre-questionnaire between experience levels (subset alpha = 0.05)

Level of training	N	Group 1 Mean	Group 2 Mean	Group 3 Mean	Group 4 Mean
PGY-1	5	40.0			
PGY-2	7		61.1		
PGY-3	3		75.0	75.0	
PGY-4	3			91.7	91.7
PGY-5	4				100.00

For each domain, a Tukey's B post-hoc analysis was conducted for the post-questionnaire scores. The knowledge domain for the post-questionnaire identified 2 distinct groups, with subsets of PGY-1,2,3 < PGY-4,5 (Table 13). The skill domain for the post-questionnaire identified 3 groups, with subsets of PGY-1,2,3 < PGY-3,4 < PGY-4,5 (Table 14). The comfort domain for the post-questionnaire resulted in 4 subset groups: PGY-1,2 < PGY-2,3 < PGY-3,4 < PGY-4,5 (Table 15).

Table 13. Post hoc (Tukey's B) analysis for Knowledge domain of post-procedure questionnaires between experience levels (subset alpha = 0.05)

Level of training	N	Group 1 Mean	Group 2 Mean
PGY-1	5	60.0	
PGY-2	7	68.6	
PGY-3	3	75.6	
PGY-4	3		95.6
PGY-5	4		100.0

Table 14. Post hoc (Tukey's B) analysis for Skill domain of post-procedure questionnaire between experience levels (subset alpha = 0.05)

Level of training	N	Group 1 Mean	Group 2 Mean	Group 3 Mean
PGY-1	5	54.7		
PGY-2	7	65.7		
PGY-3	3	75.6	75.6	
PGY-4	3		88.9	88.9
PGY-5	4			100.0

Table 15. Post hoc (Tukey's B) analysis for Comfort domain of post-procedure questionnaire between experience levels (subset alpha = 0.05)

Level of training	N	Group 1 Mean	Group 2 Mean	Group 3 Mean	Group 4 Mean
PGY-1	5	57.0			
PGY-2	7	66.4	66.4		
PGY-3	3		78.3	78.3	
PGY-4	3			95.0	95.0
PGY-5	4				100.0

Independent samples t-tests were conducted to assess the construct validity of the questionnaire between junior and senior residents for each separate domain. A significant difference ($p < 0.001$) was found for each (Table 16).

Table 16. Independent T-tests to assess construct validity of simulator
(pre = pre-procedure questionnaire, post = post-procedure questionnaire)

	Junior mean (SD)	Senior mean (SD)	*p* value	Effect size (*d*)
Knowledge Pre	58.9 (15.26)	91.3 (12.19)	**<0.001**	**2.40 (large)**
Skill Pre	53.9 (16.93)	88.7 (13.72)	**<0.001**	**2.32 (large)**
Comfort Pre	55.8 (17.43)	90.0 (12.47)	**<0.001**	**2.28 (large)**
Knowledge Post	65.0 (12.43)	91.3 (11.78)	**<0.001**	**2.19 (large)**
Skill Post	61.1 (14.45)	89.3 (11.84)	**<0.001**	**2.17 (large)**
Comfort Post	62.5 (12.34)	92.0 (10.33)	**<0.001**	**2.68 (large)**

Procedural Measurements

Descriptive analyses

Descriptive statistics were performed on the overall total scores for four measured scores: checklists, global rating scale (GRS), total errors and time to completion of the procedure (Tables 17-19). A negative skewness and steep kurtosis are noted that were skewed and kurtosed to a significant degree (as described in methods) for the virtual simulator checklist, so non-parametric tests were conducted to analyze this measure to ensure accuracy of results. The analysis of the total scores data regarding the virtual simulator was completed with one less participant than the Sawbones simulator. This participant (PGY-1, group 1) was mistakenly instructed on how to properly perform the tasks of the procedure after they had completed the Sawbones simulator and prior to competing the virtual procedure, therefore, their measured virtual data was removed from the analysis.

Table 17. Descriptive statistics of checklists and Global Rating Scales
(VR = Virtual simulator, Saw = Sawbones simulator, GRS = global rating scale, SES = standard error of skewness, SEK = standard error of kurtosis)

	N	Mean (%)	SD	Minimum Score (%)	Maximum Score (%)	Skewness (SES)	Kurtosis (SEK)
VR Checklist	21	75.9	11.73	40.0	86.7	**-1.49 (0.50)**	**3.05 (0.97)**
Saw Checklist	22	66.4	12.08	40.0	86.7	-0.22 (0.49)	-0.13 (0.95)
VR GRS	21	83.4	9.94	66.7	96.7	-0.03 (0.50)	-1.31 (0.97)
Saw GRS	22	69.0	16.62	36.7	100.0	0.24 (0.49)	-0.27 (0.95)

Table 18. Descriptive statistics of total errors during procedure
(VR = Virtual simulator, Saw = Sawbones simulator, SES = standard error of skewness, SEK = standard error of kurtosis)

	N	Mean (#)	SD	Minimum Score (#)	Maximum Score (#)	Skewness (SES)	Kurtosis (SEK)
VR Errors	21	2.6	1.20	1.0	5.0	0.23 (0.50)	-1.11 (0.97)
Saw Errors	22	4.5	1.66	2.0	8.0	0.47 (0.49)	-0.69 (0.95)

Table 19. Descriptive statistics of time to completion of the procedure
(VR = Virtual simulator, Saw = Sawbones simulator, SES = standard error of skewness, SEK = standard error of kurtosis)

	N	Mean (sec)	SD	Minimum Score (sec)	Maximum Score (sec)	Skewness (SES)	Kurtosis (SEK)
VR Time	21	641.7	120.19	469	851	0.25 (0.50)	-1.05 (0.97)
Saw Time	22	580.1	113.05	365	751	-0.31 (0.49)	-0.63 (0.95)

The residents' scores were compared between the virtual reality and Sawbones simulators. A paired samples t-test was conducted to assess for any differences between participants' performances. Table 20 demonstrates that in all cases except time to completion of the procedure, the virtual simulator resulted in residents achieving better scores. The virtual simulator scores were significantly higher than Sawbones simulator for checklist, GRS and time to completion, and significantly lower than Sawbones simulator for total errors, with medium to large effect sizes.

Table 20. Pairwise t-test of virtual and Sawbones simulator measured scores

	Virtual (SD)	Sawbones (SD)	*p* value	Effect size (*d*)
Checklist (%)	75.9 (11.73)	65.9 (12.15)	**0.017**	**0.83 (large)**
Global rating scale (%)	83.4 (9.94)	69.3 (16.99)	**0.001**	**1.05 (large)**
Total Errors	2.6 (1.20)	4.6 (1.69)	**<0.001**	**1.27 (large)**
Time to Completion (sec)	641.7 (120.19)	574.0 (112.02)	**0.033**	0.59 (medium)

Descriptive statistics for each of the PGY experience levels for each measured instrument are shown in Tables 21-24. Paired samples t-tests were conducted to assess if any significant differences existed between PGY experience levels on the two simulators. Significant differences between simulators as a function of PGY experience levels were noted for PGY-2 GRS (Table 22), PGY-2 & 3 total errors (Table 23) and PGY-5 time to completion (Table 24).

Table 21. Descriptive statistics of checklist between experience levels
(PGY-1 N = 3 for virtual, 4 for sawbones simulator)

	N	Virtual mean (%)	Virtual SD	Sawbones mean (%)	Sawbones SD	*p* value	Effect size (*d*)
PGY-1	4	63.3	19.25	58.7	13.46	0.575	0.29
PGY-2	7	75.7	8.54	63.8	13.25	0.137	1.08
PGY-3	3	84.4	3.85	72.2	13.47	0.341	1.53
PGY-4	3	80.0	6.67	72.2	8.39	0.465	1.11
PGY-5	4	79.2	6.87	71.7	12.08	0.328	0.79

Table 22. Descriptive statistics of Global Rating Scale between experience levels
(PGY-1 N = 3 for virtual, 4 for sawbones simulator)

	N	Virtual mean (%)	Virtual SD	Sawbones mean (%)	Sawbones SD	*p* value	Effect size (*d*)
PGY-1	4	71.7	4.91	53.7	13.35	0.091	2.00
PGY-2	7	82.4	7.63	61.4	6.04	**0.001**	**2.99**
PGY-3	3	92.2	7.70	71.1	9.77	0.126	2.63
PGY-4	3	94.4	1.92	76.7	10.41	0.092	2.96
PGY-5	4	82.1	9.94	94.2	7.39	0.070	1.51

Table 23. Descriptive statistics of total errors between experience levels
(PGY-1 N = 3 for virtual, 4 for sawbones simulator)

	N	Virtual mean	Virtual SD	Sawbones mean	Sawbones SD	*p* value	Effect size (*d*)
PGY-1	4	3.0	1.08	5.5	1.80	0.110	1.78
PGY-2	7	3.4	1.11	5.4	1.51	**0.032**	**1.53**
PGY-3	3	1.7	0.58	2.5	0.50	**0.038**	**1.54**
PGY-4	3	1.7	0.58	3.2	0.76	0.188	2.14
PGY-5	4	2.5	1.47	4.3	0.50	0.155	1.75

Table 24. Descriptive statistics of total time between experience levels
(PGY-1 N = 3 for virtual, 4 for sawbones simulator)

	N	Virtual mean (sec)	Virtual SD	Sawbones mean (sec)	Sawbones SD	*p* value	Effect size (*d*)
PGY-1	4	613.8	143.32	647.8	121.00	0.847	0.26
PGY-2	7	665.6	91.42	629.9	84.50	0.454	0.40
PGY-3	3	610.3	192.84	531.3	29.96	0.512	0.71
PGY-4	3	617.7	37.54	533.7	103.18	0.158	1.20
PGY-5	4	669.3	168.45	480.0	129.04	**0.018**	**1.26**

A trend was also noted with the PGY-5 residents performing worse than PGY-3/4 on virtual GRS, virtual and Sawbones total errors and virtual time.

Descriptive statistics of junior versus senior experience levels for each measured instrument are shown in Tables 25-28. Paired samples t-tests were conducted to assess if any significant differences existed between the two experience levels as a function of the virtual reality and Sawbones simulators. Significant differences between simulators were noted for junior resident GRS (Table 26), junior and senior resident total errors (Table 27) and senior resident time to completion (Table 28).

Table 25. Descriptive statistics of checklist between experience levels
(Junior N = 11 for virtual, 12 for sawbones simulator)

	N	Virtual mean (%)	Virtual SD	Sawbones mean (%)	Sawbones SD	*p* value	Effect size (*d*)
Junior	12	71.2	13.93	60.3	12.69	0.123	0.85
Senior	10	81.0	5.89	72.0	8.34	0.060	1.29

Table 26. Descriptive statistics of Global Rating Scale between experience levels
(Junior N = 11 for virtual, 12 for sawbones simulator)

	N	Virtual mean (%)	Virtual SD	Sawbones mean (%)	Sawbones SD	*p* value	Effect size (*d*)
Junior	12	78.5	8.45	57.7	10.39	**<0.001**	**2.31**
Senior	10	88.8	8.82	82.0	13.35	0.285	0.65

Table 27. Descriptive statistics of total errors between experience levels
(Junior N = 11 for virtual, 12 for sawbones simulator)

	N	Virtual mean (N)	Virtual SD	Sawbones mean (N)	Sawbones SD	*p* value	Effect size (*d*)
Junior	12	3.2	1.06	5.6	1.56	**0.004**	**1.88**
Senior	10	2.0	1.03	3.4	0.94	**0.008**	**1.40**

Table 28. Descriptive statistics of time to completion between experience levels
(Junior N = 11 for virtual, 12 for sawbones simulator)

	N	Virtual mean (sec)	Virtual SD	Sawbones mean (sec)	Sawbones SD	*p* value	Effect size (*d*)
Junior	12	646.7	108.91	630.8	98.35	0.708	0.16
Senior	10	636.1	137.32	511.5	94.08	**0.007**	**1.04**

First simulator used

The checklist, GRS, total errors and time to completion of the procedure scores were compared against the first simulator used groups. Independent samples t-tests were conducted to assess for any between score differences. For the virtual checklist, the virtual simulation first group (*M* = 70.6, SD = 12.81) had a significantly lower mean than the Sawbones first group (*M* = 81.7, SD = 7.24), t(1,21) = 2.401, $p < 0.05$, $d = 1.11$. A Mann-Whitney Test was conducted due to negative skewness and steep kurtosis of the virtual checklist. This confirmed that the virtual simulation group had a significantly lower score than the Sawbones simulation group ($p < 0.05$). No other significant differences were noted between groups.

Construct validity

Construct validity was assessed for each simulator by comparing means of the checklist scores, GRS scores, total errors and time to completion among experience levels. Two different groupings based on the experience levels of the residents were used to test whether an increase in years or level of experience resulted in better performance scores. The first group investigated all post-graduate years' (PGY 1-5) scores separately, and the second group compared junior and senior level residents.

To determine construct validity among each separate year of residency, eight one-way ANOVA's were conducted to compare scores among experience levels for each simulator (Table 29). Significant differences between groups were found for virtual GRS, Sawbones GRS and Sawbones errors.

Table 29. ANOVA to assess construct validity of simulator
(VR = virtual simulator, S = Sawbones simulator, GRS = global rating scale)

	Sum of Squares	Df	Mean Square	F	*P* value
VR Checklist (%)					
Between groups	944.05	4	236.01	2.087	0.130
Within Groups	1809.39	16	113.09		
Total	2753.44	20			
S Checklist (%)					
Between groups	660.31	4	165.08	1.167	0.360
Within Groups	2404.34	17	141.43		
Total	3064.65	21			
VR GRS (%)					
Between groups	1164.32	4	291.08	5.746	**0.005**
Within Groups	810.55	16	50.66		
Total	1974.87	20			
S GRS (%)					
Between groups	4299.98	4	1075.00	12.153	**0.000**
Within Groups	1503.68	17	88.45		
Total	5803.66	21			
VR errors					
Between groups	9.88	4	2.47	2.115	0.126
Within Groups	18.69	16	1.17		
Total	28.57	20			
S errors					
Between groups	28.61	4	7.15	4.174	**0.016**
Within Groups	29.13	17	1.71		
Total	57.74	21			
VR time (sec)					
Between groups	14834.12	4	3708.53	0.216	0.925
Within Groups	274084.55	16	17130.28		
Total	288918.67	20			
S time (sec)					
Between groups	93929.60	4	23482.40	2.288	0.102
Within Groups	174438.99	17	10261.12		
Total	268368.59	21			

Post hoc analyses were conducted for virtual GRS, Sawbones GRS and Sawbones error using Tukey's B. For virtual GRS scores: PGY-1,5,2 < PGY-5,2,3,4 (Table 30). For Sawbones simulator GRS: PGY-1,2,3 < PGY-2,3,4 < PGY-5 (Table 31). For Sawbones errors: PGY-1,2,5,4 < PGY-3,4,5 (Table 32). No distinct groups were noted.

Table 30. Post hoc (Tukey's B) analysis of virtual simulator GRS scores between experience levels (subset alpha = 0.05)

Level of training	N	Group 1 Mean	Group 2 Mean
PGY-1	4	71.7	
PGY-5	4	82.1	82.1
PGY-2	7	82.4	82.4
PGY-3	3		92.2
PGY-4	3		94.4

Table 31. Post-hoc (Tukey's B) analysis of Sawbones simulator GRS score between experience level (subset alpha = 0.05)

Level of training	N	Group 1 Mean	Group 2 Mean	Group 3 Mean
PGY-1	5	53.7		
PGY-2	7	61.4	61.4	
PGY-3	3	71.1	71.1	
PGY-4	3		76.7	
PGY-5	4			94.2

Table 32. Post hoc (Tukey's B) analysis of Sawbones total errors score between experience levels (subset alpha = 0.05)

Level of training	N	Group 1 Mean	Group 2 Mean
PGY-3	3	2.5	
PGY-4	3	3.2	3.2
PGY-5	4	4.3	4.3
PGY-2	7		5.4
PGY-1	5		5.5

A non-parametric test was conducted for the virtual checklist due to significant skewness and kurtosis. The Kruskall Wallis demonstrated no significant difference between the scores of training levels for the virtual checklist (p = 0.273).

Eight independent t-tests were conducted to differentiate between the scores of the four measured residents' performances on each simulator (checklist, GRS, total errors and time to completion of the procedure) of junior and senior residents. Senior residents scores were significantly better than junior residents in all scored results, except for the virtual simulator checklist and time to completion (Table 33). The effect size differences in each of these measurements are large (>0.8).

Table 33. Independent t-tests to assess construct validity of simulators (junior versus senior) (VR =virtual simulator, S = Sawbones simulator, GRS = global rating scale)

	Junior mean (SD)	Senior mean (SD)	*p* value	Effect size (*d*)
VR Checklist (%)	71.2 (13.92)	81.0 (5.89)	0.054	0.93
S Checklist (%)	61.7 (12.99)	72.0 (8.34)	**0.043**	**0.98 (large)**
VR GRS (%)	78.5 (8.45)	88.8 (8.82)	**0.013**	**1.20 (large)**
S GRS (%)	58.2 (10.04)	82.0 (13.35)	**<0.001**	**1.98 (large)**
VR errors	3.2 (1.06)	2.0 (1.03)	**0.014**	**1.18 (large)**
S errors	5.5 (1.56)	3.4 (0.94)	**0.002**	**1.65 (large)**
VR time (sec)	646.7 (108.96)	636.1 (137.34)	0.846	0.09
S time (sec)	637.3 (96.46)	511.1 (94.08)	**0.006**	**1.32 (large)**

A multivariate analysis of variance (MANOVA) was also conducted between the two simulators. The virtual checklist score of the senior residents (M = 81.0, SD = 5.89) was found to be significantly higher than the junior residents (M = 71.2, SD = 13.92), $F(1,21) = 5.115$, $p < 0.05$, $d = 0.93$. A significant difference in scores was confirmed in each of the results shown in Table 33.

Comparisons between first simulator used groups and experience levels are shown in Figures 3-10 for each measure. Each figure demonstrates the senior residents doing significantly better than junior residents except for virtual time.

Figure 3. Virtual simulator checklist scores

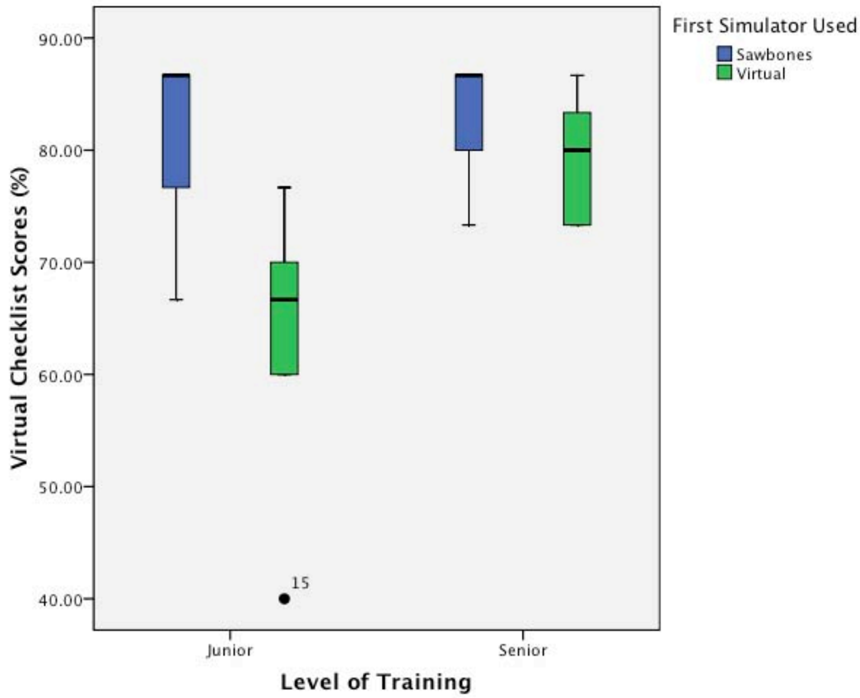


Figure 4. Sawbones simulator checklist scores

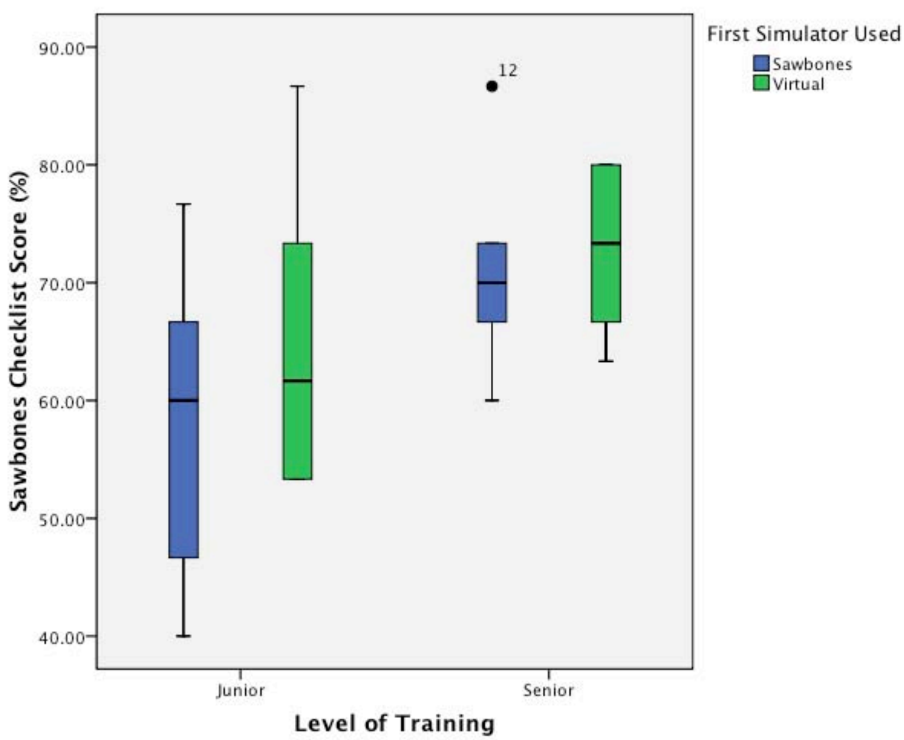


Figure 5. Virtual simulator Global Rating Scale (GRS) scores

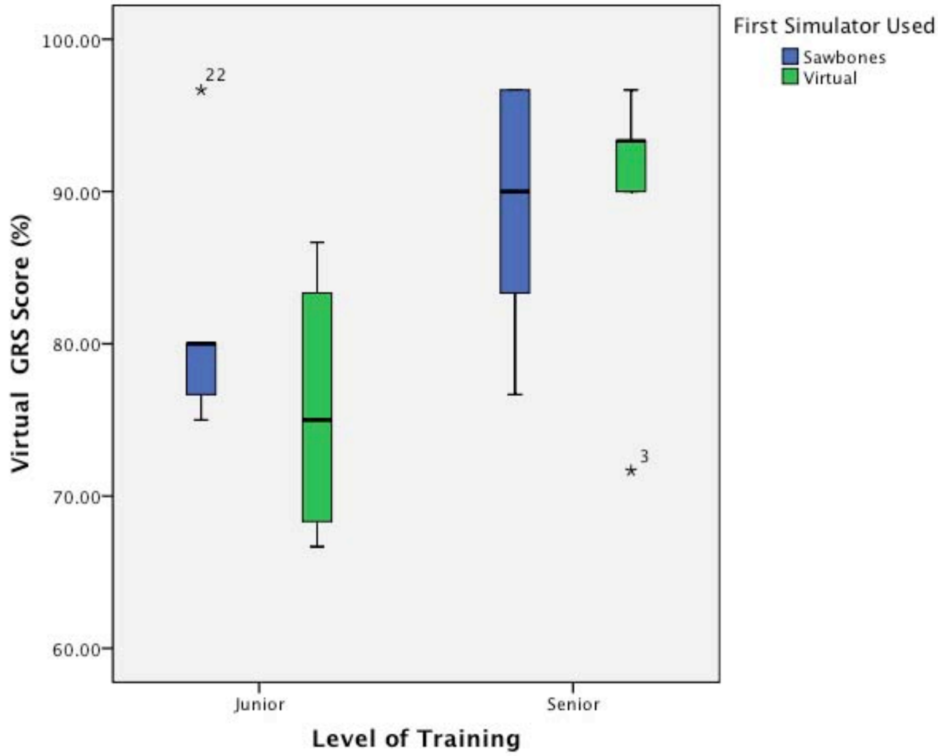


Figure 6. Sawbones simulator Global Rating Scale (GRS) scores

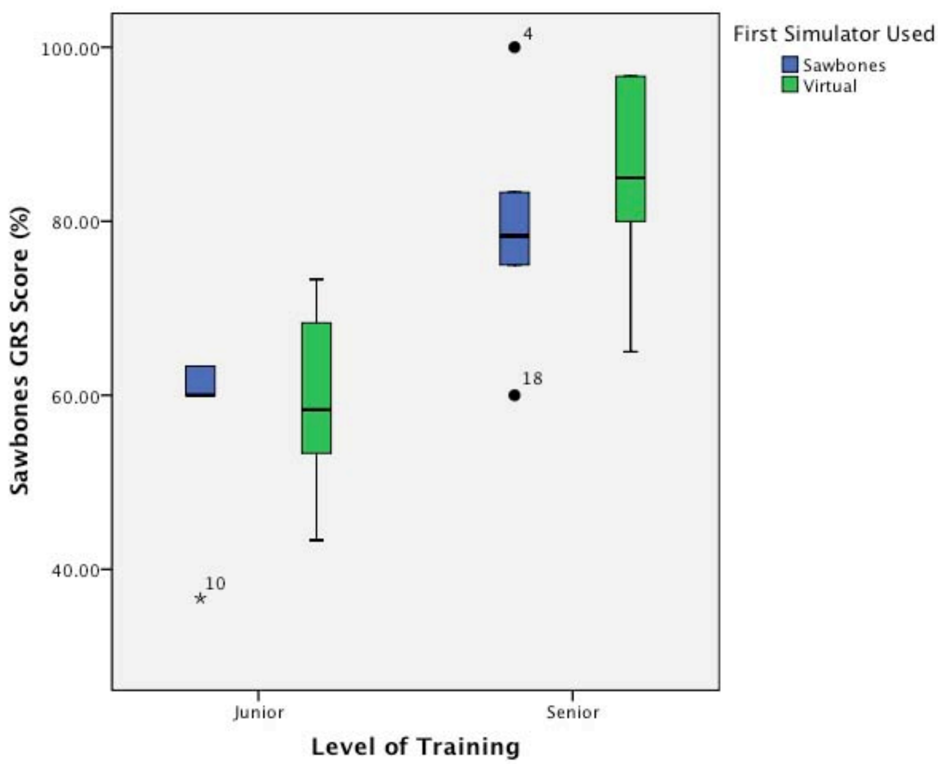


Figure 7. Virtual simulator total error scores

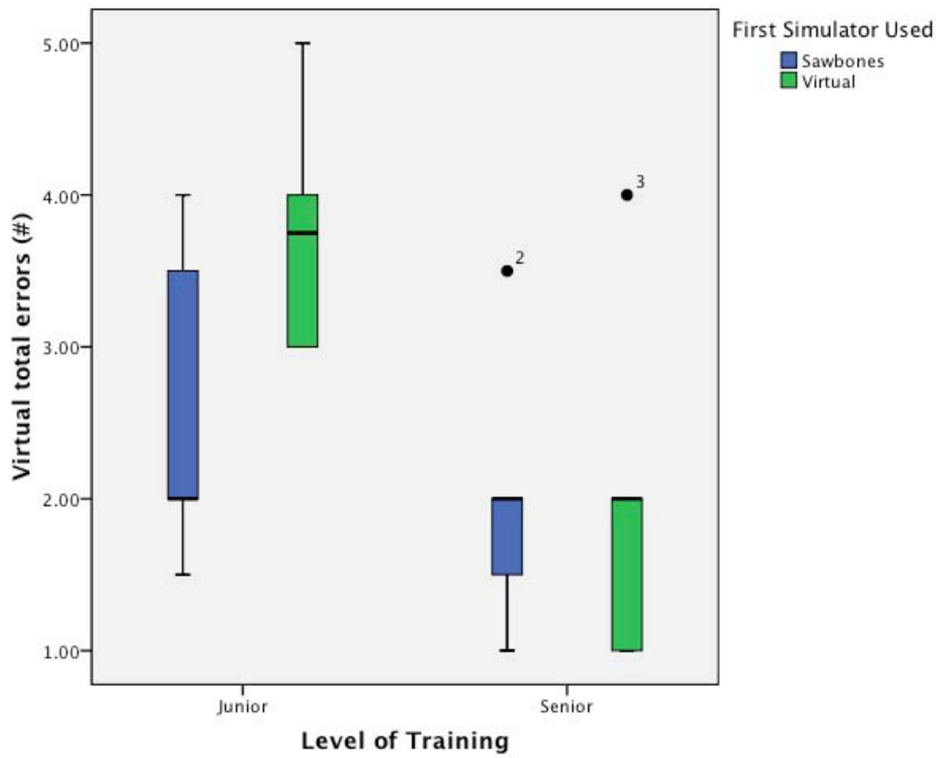


Figure 8. Sawbones total error scores

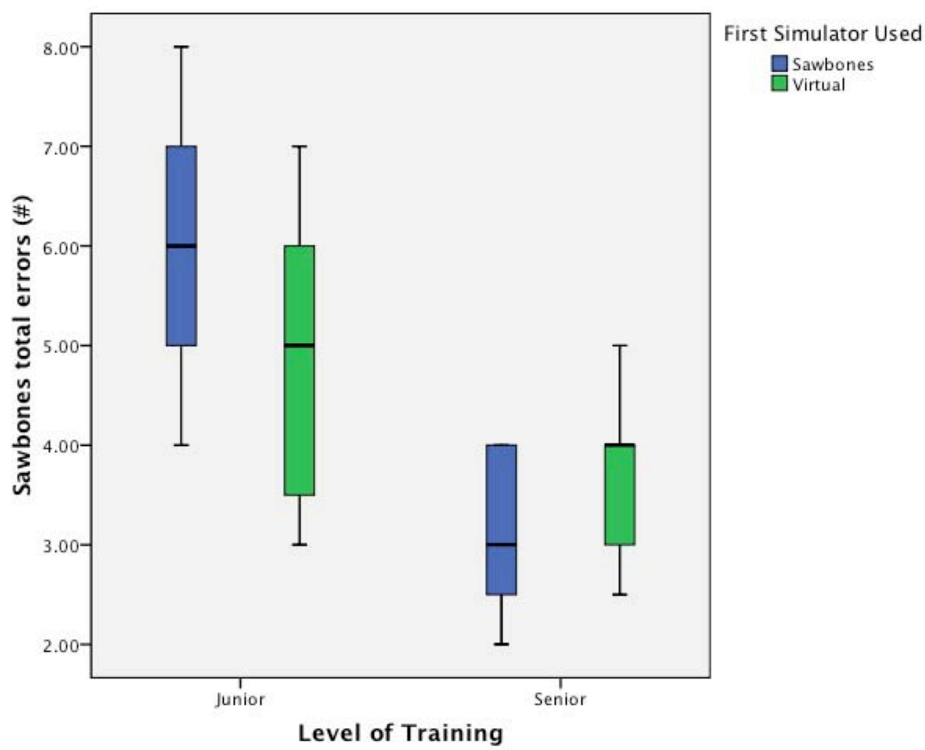


Figure 9. Virtual simulator time to completion of the procedure

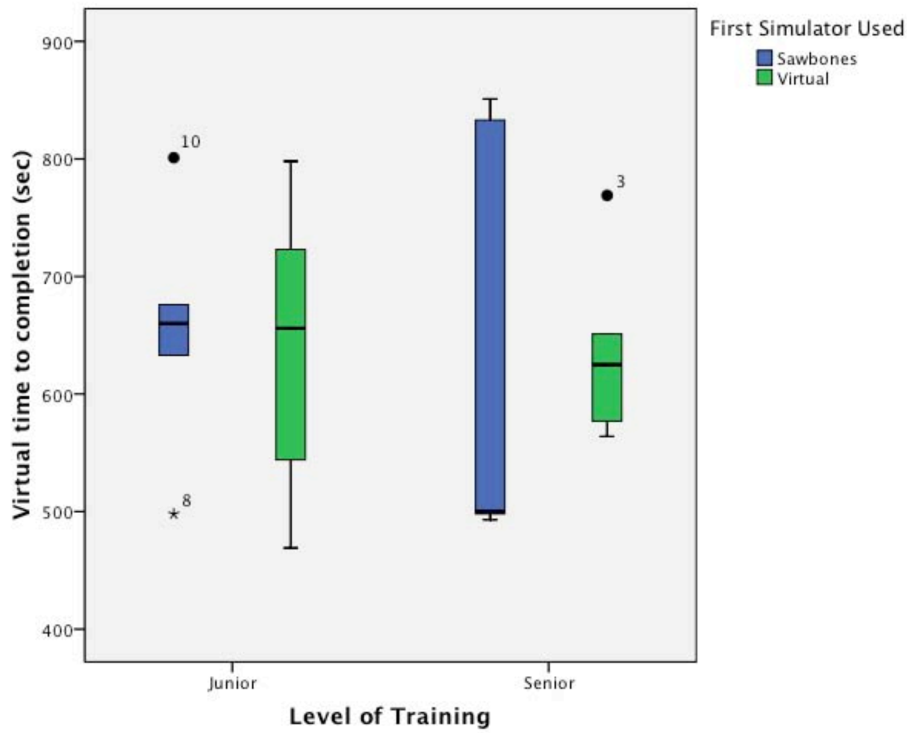
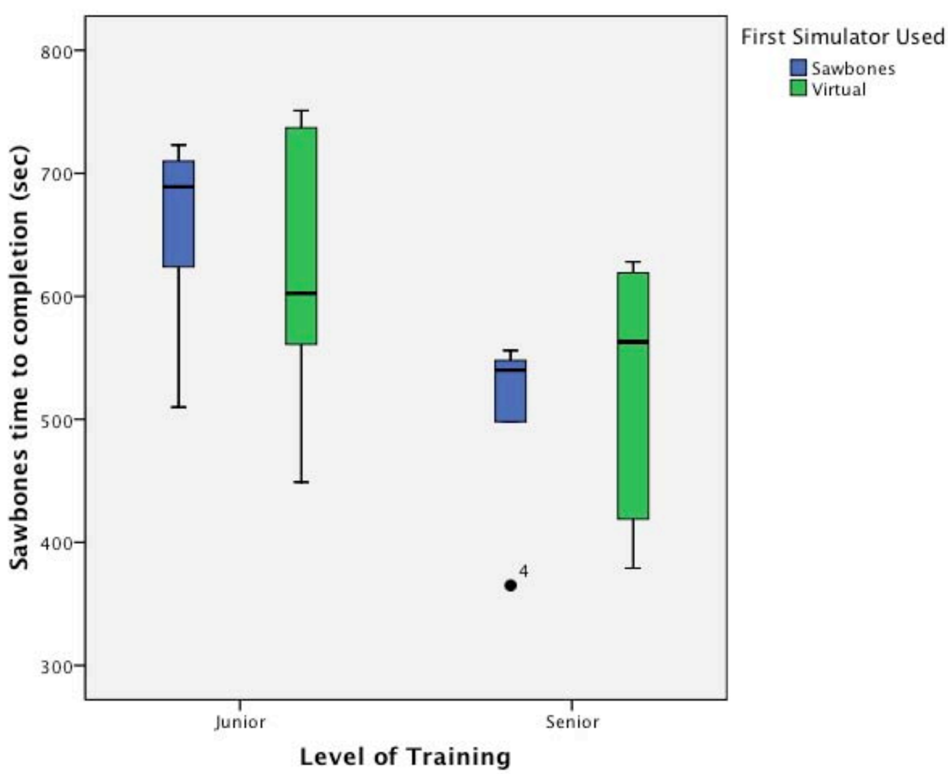


Figure 10. Sawbones simulator time to completion of the procedure



Since the virtual checklist data were skewed and kurtosed to a significant degree, a Mann-Whitney test was conducted. No significant difference between junior or senior residents' scores were found (p = 0.072)

Criterion validity

The criterion validity was evaluated using a Pearson's product moment correlation coefficient (r) within and between simulators. The within simulators analysis yielded significant Pearson's product moment correlation's (Table 34). Time was not significantly correlated with any score, except in the case of Sawbones time to GRS.

Table 34. Pearson's product moment correlation coefficient (*r*) within simulators

	Checklist to GRS	Checklist to Error	GRS to Error	GRS to Time
Virtual	0.69 ***	-0.48*	-0.54*	-0.32
Sawbones	0.57**	-0.81***	-0.58**	-0.65***

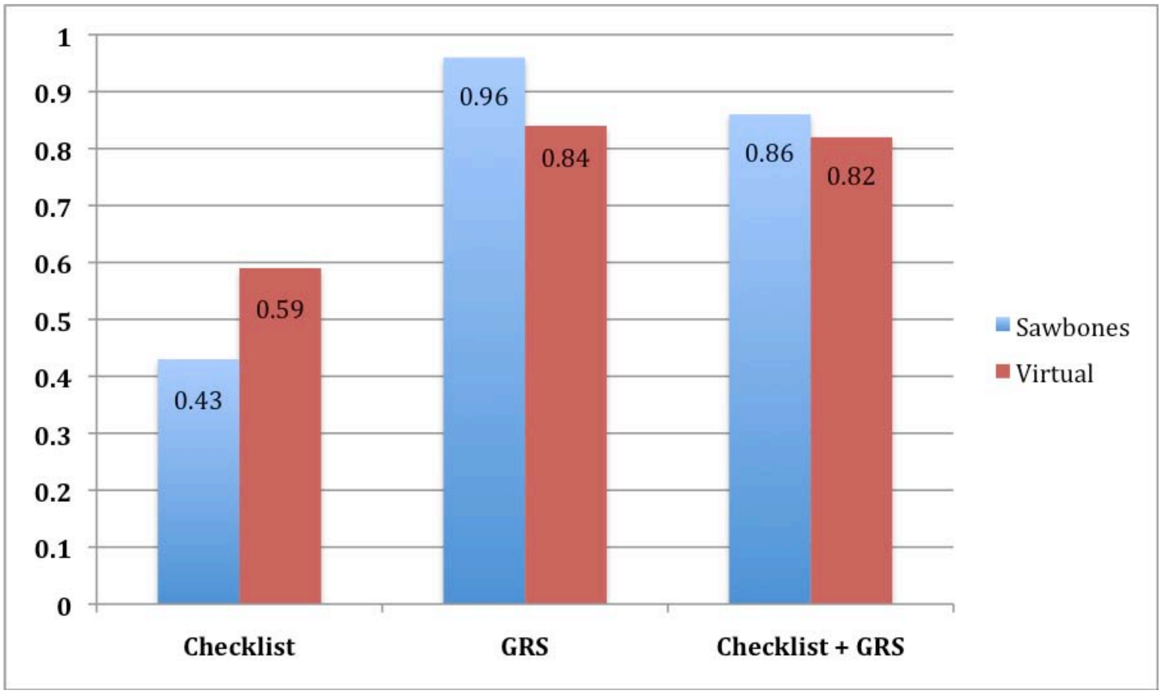
(* p<0.05, ** p<0.01, ***p<0.001)

There were no significant Pearson's coefficients on any of the measures when a correlation analysis was conducted between the two simulators.

Internal consistency

The internal consistency of the checklists and GRS were determined by using Cronbach's alpha (α). Neither the virtual (α = 0.59) nor Sawbones (α = 0.43) simulator checklists' reached the acceptable level of α = 0.70 established for this study. When the checklist was combined with the GRS, the total internal consistency reached a gold standard level of reliability established at or above 0.80 (Figure 11).

Figure 11. Internal Consistency of Checklist and GRS (Cronbach's α)
(GRS =global rating scale)



Inter-rater reliability

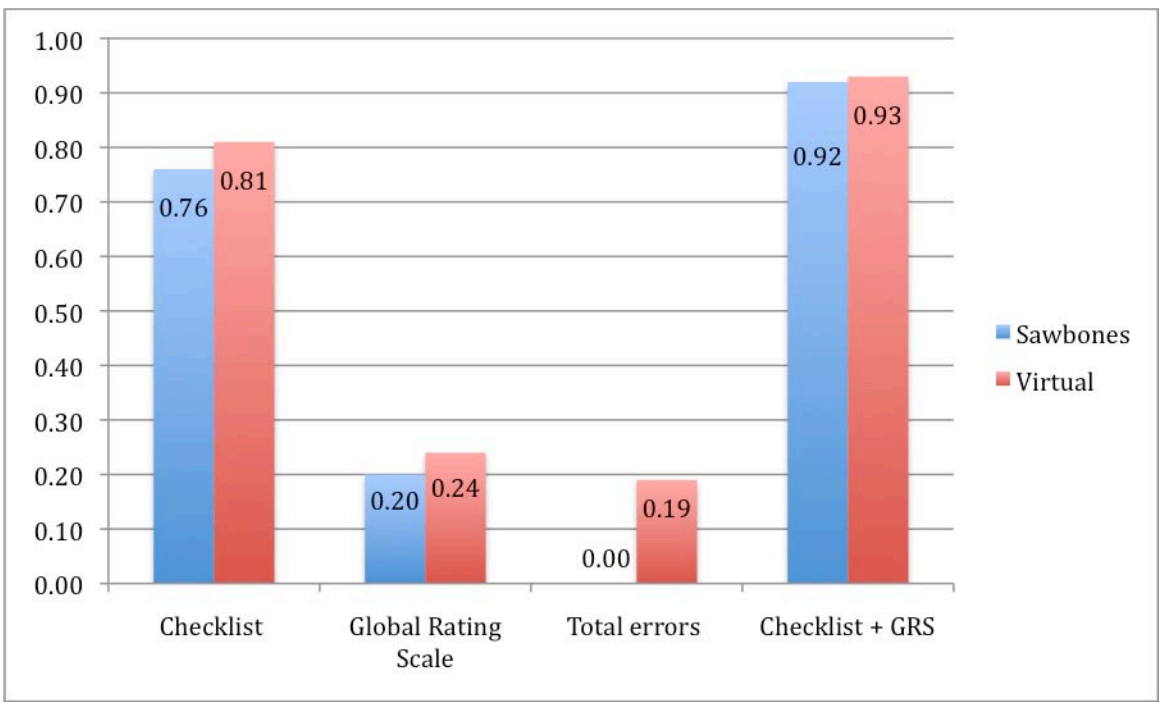
The inter-rater reliability was assessed by calculating a Cohen's Kappa coefficient for the checklists, global rating scales, and total errors for each simulator. Two independent examiners evaluated ten participants for the virtual simulator and six participants for the Sawbones simulator. The Cohen's Kappa coefficient was calculated for the checklists and Global Rating Scales (Table 35). Significant inter-rater reliability was demonstrated in checklists alone, and in combination with GRS.

Table 35. Kappa (*k*) coefficient for checklists and Global Rating Scales

	Kappa (*k*)	*p* value
Virtual checklist	0.81	**<0.001**
Sawbones checklist	0.78	**<0.001**
Virtual GRS	0.16	0.036
Sawbones GRS	0.13	0.157
Virtual checklist + GRS	0.67	**<0.001**
Sawbones checklist + GRS	0.65	**<0.001**

The Kappa coefficient for the total errors could not be calculated due to missing data points, so an intraclass coefficient was calculated for all scores once again (Figure 12). Both checklists demonstrated significant inter-rater reliability ($p < 0.001$) above 0.70, and when the checklist and GRS were combined, both reach an intraclass coefficient above 0.90 ($P < 0.001$).

Figure 12. Inter-rater reliability of checklist, global rating scale (GRS) and total errors (using intraclass coefficient)



Post-procedure simulator-specific Questionnaires

Descriptive analysis

Post-procedure questionnaire data from the two residents who participated in the pilot project were added to the full study data as neither the simulators nor the post procedure questionnaires were changed between these dates; descriptive statistics are shown in Table 36. A negative skewness and steep kurtosis are noted that were skewed and kurtosed to a significant degree (as described in methods) for the Sawbones environment, so non-parametric tests were used to assess this parameter.

Table 36. Descriptive statistics of Questionnaires between domains
(VR = Virtual simulator, Saw = Sawbones simulator, SES = standard error of skewness, SEK = standard error of kurtosis, Enviro = environment domain, Equip = equipment domain, Psych = psychological domain)

	N	Mean (%)	SD	Minimum Score (%)	Maximum Score (%)	Skewness (SES)	Kurtosis (SEK)
VR Enviro	24	68.3	10.21	48.0	84.0	-0.39 (0.47)	-0.64 (0.92)
Saw Enviro	24	82.5	10.40	52.0	100.0	**-0.99** (0.47)	**2.27** (0.92)
VR Equip	24	60.4	13.56	36.7	86.7	0.50 (0.47)	-0.27 (0.92)
Saw Equip	24	81.0	10.38	56.7	96.7	-0.77 (0.47)	0.25 (0.92)
VR Psych	24	50.3	11.92	27.5	72.5	0.01 (0.47)	-0.31 (0.92)
Saw Psych	24	71.9	10.33	52.5	95.0	0.30 (0.47)	0.52 (0.92)

Fidelity

The post-procedure simulator-specific questionnaires were used to assess three separate domains of fidelity: environment, equipment, and psychological. The separate domains were compared between both first simulator used groups and experience level groups (PGY's separately, and junior versus senior).

An independent samples t-test compared the scores for each domain of fidelity for first simulator used. No significant difference in the scores between group 1 and group 2 were found. This finding was corroborated when a one-way ANOVA and MANOVA were run for these variables.

In order to determine if the means for the domains of fidelity were different among levels of experience (PGY's separately), six one way ANOVA's were conducted. The virtual simulator environment domain yielded a significant difference between groups, $F(4,24) = 3.90$, $p < 0.05$. A post hoc analysis was conducted using Tukey B, but no distinct groups were identified, as noted in Table 37. It appears, however, that the PGY-3 and 4's gave significantly lower mean scores than PGY-1,2,5 for the virtual simulator environment domain.

Table 37. Post hoc (Tukey's B) analysis for virtual Environment between experience levels (subset alpha = 0.05)

Level of training	N	Group 1 Mean (SD)
PGY-4	3	58.7 (10.07)
PGY-3	5	59.2 (8.67)
PGY-2	7	72.0 (6.93)
PGY-5	4	72.0 (11.31)
PGY-1	5	75.2 (5.93)

Equal variance was not found for the Sawbones environment or equipment domains, therefore a Kruskall Wallis test was conducted to see if there were any differences in residents' performance between training levels. A significant difference between scores was found for the Sawbones environment; $\chi^2(4,24) = 10.915$, $p < 0.05$. This test identified that PGY-2,3,4 (7.93, 10.10, 10.67) ranked the Sawbones environment lower than PGY 1,5 (15.50, 21.13) residents.

A MANOVA was conducted with the three fidelity domains for each simulator being the dependent variables, and the simulator first used group and experience level groups as the independent variables. No significant differences among the means between groups (first simulator used and PGY's separately) was found, with the virtual environment having a $p = 0.073$.

To compare the scores of fidelity domains among the junior and senior resident groups, independent samples t-tests were conducted. The junior residents ($M = 73.3$, SD = 6.46) provided the virtual environment with a significantly higher score than the senior residents ($M = 63.3$, SD = 11.03), $t(1,24) = 2.71$, $p < 0.05$, $d = 1.11$. No other domains showed any significant difference between group scores. An ANOVA and MANOVA were also conducted, and again a significant difference among experience levels of junior and senior residents existed only for the virtual environment domain. Non-parametric tests were conducted for the Sawbones environment domain, as significant negative skewness and steep kurtosis were noted. No significant differences were found.

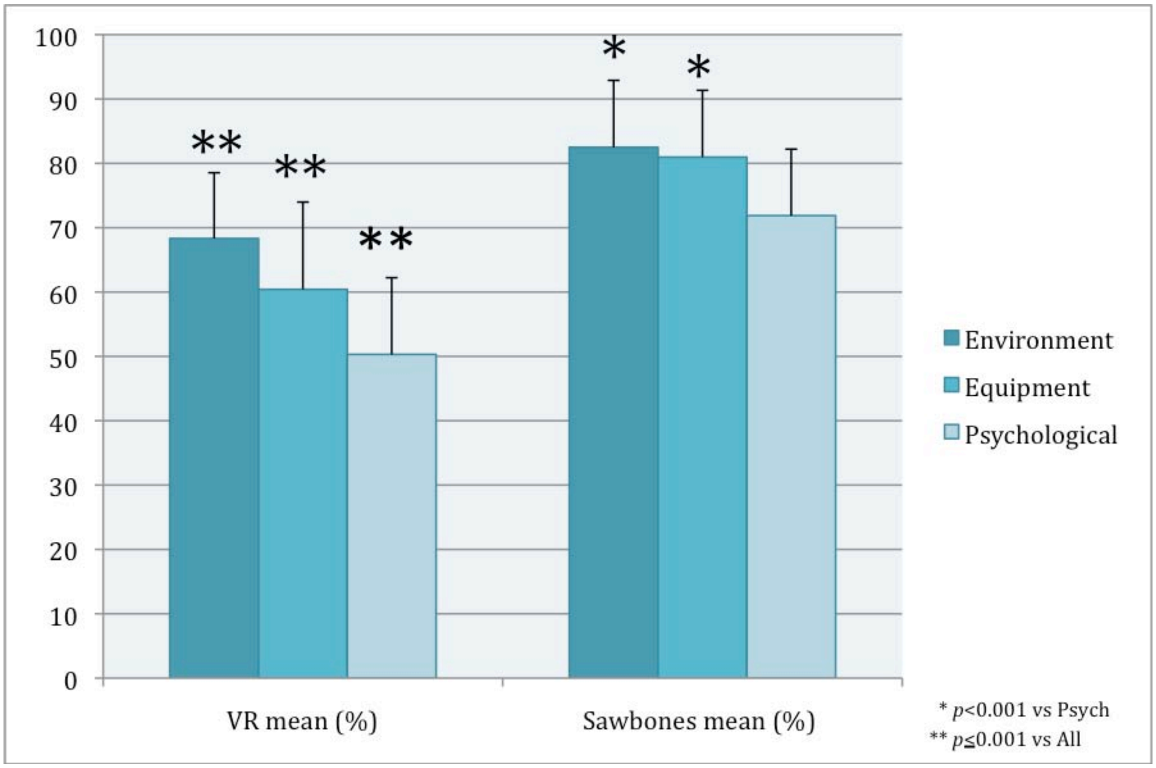
The scores of the three fidelity domains were compared between the two simulators using three paired samples t-tests. In all three domains, the mean percentage scores of the Sawbones simulator were higher than the virtual simulator (Table 38).

Table 38. Pairwise comparison of fidelity domains among simulators (%)

	Virtual mean (SD)	Sawbones mean (SD)	*p* value	Effect size (*d*)
Environment	68.3 (10.21)	82.5 (10.40)	**<0.001**	**1.38**
Equipment	60.4 (13.56)	81.0 (10.38	**<0.001**	**1.71**
Psychological	50.3 (11.92)	71.9 (10.33)	**<0.001**	**1.96**
Overall Fidelity	59.7 (10.20)	78.5 (8.81)	**<0.001**	**2.31**

Paired samples t-tests were conducted to assess if each individual domain held significantly different scores than the others for each separate simulator. The virtual simulator demonstrated significantly different mean scores among all fidelity domains, $p \leq 0.001$ (Figure 13). The Sawbones simulator mean scores among fidelity domains were not significantly different between the environment and equipment scales, but both were significantly different from the psychological domain scale, $p < 0.001$ (Figure 13).

Figure 13. Comparison of means between fidelity domains for each simulator (VR = Virtual Simulator, Psych = Psychological)



To assess the internal consistency of the questionnaire domains, a Cronbach's α was calculated for each separate domain, and all together (overall fidelity). The internal consistency of the questionnaire domains are reported in Table 39, with all domains reaching at least 0.70, except for the virtual environment.

Table 39. Internal Consistency (Cronbach's α) of Questionnaire Domains

	Environment	Equipment	Psychological	Overall Fidelity
Virtual	0.56	0.78	0.83	0.88
Sawbones	0.74	0.76	0.78	0.89

Table 40 provides correlation coefficients between each fidelity domain (i.e., environment, equipment and psychological) subscale on the questionnaire, as well as the overall fidelity scores. Significant correlations ($p < 0.05$) were identified in all bivariate correlations except between the virtual environment and Sawbones equipment subscale scores. Each domain correlated significantly with all other domains of the same simulator, and with its specific fidelity domain between the two simulators. The overall total fidelity score of the virtual simulator correlated with the Sawbones simulator at $r = 0.71$.

Table 40. Pearson's coefficient (*r*) for fidelity domains within and between simulators (VR = virtual simulator, Saw = Sawbones simulator, Enviro = environment domain, Equip = equipment domain, Psych = psychological domain, fidelity = overall fidelity)

	Virtual Enviro	Sawbones Enviro	Virtual Equip	Sawbones Equip	Virtual Psych	Sawbones Psych	Virtual Fidelity
VR Enviro	1						
Saw Enviro	**0.52****	1					
VR Equip	**0.65****	**0.44***	1				
Saw Equip	0.39	**0.58****	**0.56****	1			
VR Psych	**0.62****	**0.48***	**0.54****	**0.44***	1		
Saw Psych	**0.56****	**0.55****	**0.52***	**0.63****	**0.64****	1	
VR Fidelity	**0.84*****	**0.56****	**0.85*****	**0.60****	**0.88*****	**0.67*****	1
Saw Fidelity	**0.58****	**0.79*****	**0.60****	**0.86*****	**0.63****	**0.89*****	**0.71*****

(* p<0.05, ** p<0.01, ***p<0.001)

A paired samples t-test was conducted to compare the mean scores of the three fidelity scales between the Sawbones and virtual simulators (Table 41). The questionnaires used a 5-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). A significant difference between simulators (Sawbones versus virtual) was found in 3/5 questions for environment, 5/6 for environment, and 8/8 for psychological domains, all yielding large effect sizes differences. Questions regarding learner satisfaction were also analyzed (questions 20-24), in which 4/5 questions yielded a significant difference between simulators, all with medium and large effect sizes.

Table 41. Descriptive statistics for each question

Question	Sawbones Mean	Sawbones SD	Virtual Mean	Virtual SD	*p* value	Effect Size (*d*)
Environment						
1) This simulator was responsive to the actions performed (overall)	4.3	0.44	3.6	0.65	**<0.001**	**1.14**
2) The controls were not problematic to use in this simulator	4.3	0.82	3.1	0.85	**<0.001**	**1.44**
3) Visual representation of the forearm was realistic enough for the procedure	3.8	0.93	3.6	0.93	0.31	0.23
4) Visual representation of the tools in this simulator are important in the performance of this procedure	4.3	0.75	4.2	0.88	0.69	0.10
5) The general performance using this simulator was close in comparison to my general performance in the clinical settings	4.0	0.69	2.5	0.78	**<0.001**	**1.93**
Equipment						
6) This simulator demonstrated precise movements of tools	4.2	0.57	3.1	1.15	**<0.001**	**1.21**
7) All tools/equipment required were accessible during the procedure	4.5	0.72	4.0	0.89	0.06	*0.57*
8) Tactile force feedback was simulated accurately on this simulator	3.8	0.76	2.7	0.86	**<0.001**	**1.38**

Table 41 continued. Descriptive statistics for each question

Question	Sawbones Mean	Sawbones SD	Virtual Mean	Virtual SD	*p* value	Effect Size (*d*)
9) Placement of tools was properly simulated on this simulator	4.3	0.62	3.2	0.87	**<0.001**	**1.50**
10) Drilling through bone was accurate on this simulator	3.8	0.85	2.5	0.83	**<0.001**	**1.44**
11) Plunging (exiting second cortex) of the drill was easy to feel on this simulator	3.8	1.02	2.6	1.21	**<0.001**	**1.09**
Psychological						
12) While performing this procedure on this simulator, it felt like I was actually doing the procedure on a patient	3.3	0.85	2.5	0.83	**<0.001**	**0.94**
13) I felt comfortable performing the procedure	4.5	0.59	3.7	1.05	**<0.001**	**1.01**
14) I felt like all my senses (not sound on VR) were engaged during the procedure	4.3	0.85	3.4	0.93	**0.001**	**0.93**
15) The actual drilling made me feel as though I were performing a real procedure (in OR)	3.6	0.82	2.1	0.93	**<0.001**	**1.77**
16) The visual aspects of the environment (i.e. Display, Haptics device, table) made me feel as if I were performing the real procedure (in OR)	3.4	0.82	2.2	0.93	**<0.001**	**1.34**
17) The feel of the equipment made me feel as if I were actually doing the real procedure (in OR)	3.8	0.76	2.2	0.82	**<0.001**	**2.10**
18) The events around me made me feel as though I were actually doing the real procedure (in OR)	2.7	1.01	1.8	0.72	**0.001**	**1.02**
19) My experience in the simulator's environment (overall) seemed consistent with my real world experiences	3.3	0.85	2.3	0.82	**<0.001**	**1.10**
Learner satisfaction						
20) This simulator is an effective method for learning basic surgical fixation procedures	4.3	0.55	3.7	0.87	**0.001**	**0.87**
21) This simulator is effective for the **introduction** of basic surgical skills	4.6	0.50	4.2	0.93	0.083	*0.52*

Table 41 continued. Descriptive statistics for each question

Question	Sawbones *Mean*	Sawbones SD	Virtual *Mean*	Virtual SD	*p* value	Effect Size (*d*)
22) This simulator is an effective method to practice **previously learned** techniques for my surgical training	3.7	0.91	2.6	1.14	**<0.001**	**1.10**
23) Simulator-specific based examinations would be useful for the assessment of surgical fixation of the ulna	3.7	0.87	3.0	0.89	**0.04**	*0.76*
24) This simulator would be valuable for refresher skills	3.7	0.91	3.0	0.98	**<0.001**	*0.75*
Miscellaneous						
25) People were available to answer my questions when needed during the procedure	4.5	0.51	4.8	0.53	**0.031**	0.48
26) This simulator (overall) provided a challenging surgical experience	2.8	0.99	3.5	0.83	**<0.001**	*0.78*
27) Further development of the simulator is needed prior to it being a formal evaluation tool	2.8	0.87	4.6	0.50	**<0.001**	**2.63**
28) Prior simulation experience is needed prior to examination using this simulator	2.6	1.02	2.8	1.38	0.540	0.18
29) I would likely use this simulator in my spare time for practicing procedures, if it were readily available	3.6	1.14	3.8	1.03	0.575	0.16
30) I would be more likely to use this simulator in my spare time than the other simulator for practicing procedures, if both were readily available	3.4	1.08	2.2	0.84	**0.009**	**1.30**
31) Fracture fixation using this simulator was a valuable experience	3.7	0.81	3.4	0.97	0.166	0.33
32) This simulator should be included in residency training program	4.1	0.72	3.3	1.19	**0.001**	**0.87**
33) A surgical skills laboratory would be valuable to my surgical training	4.8	0.42	4.8	0.61	0.575	0.08

As part of the simulator specific questionnaires, five short answer questions were completed by the participants regarding the specific simulator used. They included 1) the simulators' strengths; 2) the simulators' weaknesses; 3) the frustrations experienced with that simulator; 4) what the participants would change on the simulator to make it better; and 5) what the benefits would be of using this type of simulator for surgical education? Participants' responses are summarized together in Table 42. Numbers in parentheses are the number of participants who responded this way (out of 24).

Table 42. Short answers for both the virtual and Sawbones simulators
( ) = # of respondents

	Virtual Simulator	Sawbones Simulator
Strengths of this simulator?	**Good for learning and reviewing basic steps of the procedure and equipment** (16, 67%) **Good graphics** (6, 25%) Tools easily available (3, 13%) 6 degrees of freedom with force feedback (2, 8%) Adaptable for left-handed (2, 8%)	Use real tools (18, 75%) Immediate tactile feedback (11, 46%) **Visually realistic ulna/tools (7, 29%)** **Good for learning and reviewing basic steps of the procedure and equipment (4, 17%)** Has sound (2, 8%) Using both hands (1, 4%)
Weaknesses of this simulator?	**Tactile force not correct** for: feeling cortex (drill, tap), screwing, "artifact felt while drilling" (20, 83%) **Not realistic enough** for: instrument handle, **surgical approach**, and transparency, "click" reduction, depth gauge, visuals (9, 38%) No sound (5, 21%) **No fracture to reduce** (3, 13%) Not representative of OR (3, 13%) Only using one hand (2, 8%) No stereoscopic vision (1, 4%) **"Too easy"** (1, 4%)	**No approach or fracture to reduce (16, 67%)** Sawbones **rubber/thin (osteoporotic) feel (10, 42%)** **Not like real surgery**: no stress/anxiety, has laboratory feel, **no surgical approach (3, 13%)** Messy (3, 13%) Only one bone (2, 8%) **"Too easy"** (1, 4%)

Table 42 continued. Short answers for both the virtual and Sawbones simulators
( ) = # of respondents

Frustrations of this simulator?	**Tactile force not correct** (as above) (11, 46%) Drill did not go straight through bone (rose upwards) (5, 21%) Depth difficult to assess (4, 17%) "Learning curve" with simulator (4, 17%) "snap points" not realistic (2, 8%) Inaccuracy of tool placement (i.e. drill in drill guide) (2, 8%) Cannot assess reduction/screw placement (1, 4%) Difficult to manipulate/target (1, 4%)	**Not like real bone** (7, 29%) No fracture to reduce (3, 13%) No soft tissue (1, 4%) No approach (1, 4%)
What would you change?	More realistic screw/drilling/ tapping (17, 71%) Make haptics handle same as tool (6, 25%) Add sound (3, 13%) **Perform reduction, lag screw** (3, 13%) **Surgical dissection** and/or complications (3, 13%) Two haptics devices (2, 8%) Correlate visuals and haptics (drilling) (2, 8%) Add stereoscopic vision (1, 4%) **Add gloves/gown/mask** (1, 4%)	Add fracture to reduce, **Perform reduction, lag screw (12, 50%)** Have both radius and ulna (2, 8%) **Surgical dissection (2, 8%)** Have denser bone (1, 4%) Have an assist to hand instruments (as in real OR) (1, 4%) **Add gloves/gown/mask (1, 4%)**
Benefits of this type simulator?	**Practice** steps, learning sequence, introduction of skills for early learners (13, 54%) Quick, clean, **safe** (no patient) (8, 33%) Low stress environment (2, 8%) Can produce complex procedures, fractures, scenarios (2, 8%) **May use in skills laboratory** (1, 4%)	Tools same as OR (11, 46%) Best for inexperienced and junior residents (8, 33%) Get to know implants, instruments (5, 21%) **Practice** procedure, reduction (3, 13%) **May use in skills laboratory (1, 4%)** Cleaner, **safer** than cadavers (1, 4%) Minimal learning curve (1, 4%)

Feasibility

The cost associated with creating a virtual simulator and the costs of using Sawbones were compared. The initial costs of a virtual simulator for fracture fixation would cost approximately $85,000.00, and to run it for 5 years would cost approximately $105,000.00 overall (Table 43). The added costs that would occur are due to software upgrades and personnel time installing this software and fixing problems, which are estimated over the five years of repeated use of the simulator.

Table 43. Approximate initial and 5 year costs of virtual simulator

	Initial Costs	Over 5 years
Computer and Display	$5,000.00	$0
Haptics Device	$60,000.00	$10,000.00 (upgrades)
Software	Included in Haptics	$0
Personnel (student)	$20,000.00	$10,000.00
Total costs	**$85,000.00**	**$20,000.00**

The initial costs of a fracture fixation model using Sawbones are less since the tools and models have been previously developed. To set up a Sawbones simulator for surgical fixation of an ulna with all its tools, the cost would be approximately $3,445.00 without the Sawbones (Table 44). To run this simulator over a 5-year period would cost approximately $16,150.00 without the actual Sawbones, as drill bits would dull and certain instruments would break or need replacement. If each resident (25) were to use 10 Sawbones per year for practice with this simulator, over 5 years the total cost would be approximately $55,150.00.

Table 44. Approximate initial and 5 year costs of Sawbones simulator

	Initial Costs	Over 5 years
Sawbones (ulna)	$30.00 each	$30/person/time
Plates	$100.00 each	$100 x 10 = $1,000.00
Cortex Screws	$20.00 each x 6 = $120.00	$20 x 120 = $2,400.00
Drill bits	$55 each x 2 = $110.00	$55 x 100 = $5,500.00
Drill guides	$400.00 x 2 = $800.00	$400 x 4 = $1,600.00
Taps	$65.00	$65 x 10 = 650.00
T-handle (tap)	$250.00	$250 x 2 = $500.00
Drill	$150.00	$150 x 2 = $300.00
Base/Vice	$300.00 x 2 = $600.00	$300 x 4 = $1,200.00
Depth Gauge	$300.00	$300 x 3 = $900.00
Reduction forceps	$250.00 x 3 =$750.00	$250 x 6 = $1,500.00
Screwdriver	$200.00	$200 x 3 = $600.00
Total costs (no Sawbones) **Total** (with # Sawbones/yr/resident)	$3,445.00 (+ 1 Sawbones each =) **$4,225.00**	$16,150.00 (+ 10 each/year =) **$55,150.00**

Another option is renting the needed equipment, which would cost approximately $2000 for a 4 person, 4-hour session once. To perform twelve of these sessions a year (enough for each resident to attend twice), the cost would be $24,000.00. To rent this simulator for 5 years, with each resident only practicing twice per year, the cost would be $120,000.00.

Exit interviews were conducted for both students participating in the creation of this virtual simulator. Once their portion of the development for the virtual simulator was complete, they were immediately asked each question. Their answers in regards to the development of this simulator are grouped in Table 45.

Table 45. Exit interview answers

Questions	Answers
Background	One 2nd year engineering summer student One 1st year Master's of engineering student Neither with prior experience developing virtual simulators or code
Aid readily available?	A PhD engineering student and their Supervisor, both with experience in virtual modeling, were easily available
Total time developing model?	30 hrs/week x four months learning code and developing basic simulator 15 hrs/week x four months building major parts of simulator 7hrs/week x two months doing minor changes and upgrades Approximately 776 hrs spent learning code and developing simulator
Satisfaction with simulator?	Thought it was a fair simulator, though for training purposes required more accuracy Felt if they had prior knowledge of virtual environment and code, and more time, simulator would have been better
Easiest part of development?	Defining and dealing with graphical objects and visualization
Hardest part of development?	Rendering the force feedback sensations
Frustrations during development?	Feeling rushed (timeline) Implementation of rendering forces – which led to reduced reality of force feedback

CHAPTER 5 – DISCUSSION

This chapter will discuss the findings in an attempt to answer the four proposed research questions:

1) Does surgical fixation of an ulna fracture on a virtual bone simulator have the same or comparable fidelity at all levels of resident training as performing the same subtasks on Sawbones?
2) Can we measure participant performance and quality of surgical skills using a modified checklist and global rating scale format?
3) Can we accurately assess the learners' perceptions of their experiences related to specific subtasks of the procedure using a questionnaire?
4) Is it feasible to develop a virtual fracture fixation model for orthopaedic surgery residents to perform surgical fixation of an ulna.

The limitations of the study, conclusion, and future directions will close the chapter.

Question 1: Fidelity

High fidelity simulators have been used for decades in the aviation industry to train pilots prior to actually flying in hopes of preventing loss of resources, including trainees, trainers, and planes [24]. Simulators are beginning to be used in the medical fields, to aid in preventing loss of operating time and avoiding injuries to patients. The fidelities of both the virtual and Sawbones simulators were assessed with the post-procedure simulator specific questionnaires. No previous literature has been found that evaluates the fidelity of the

Sawbones simulator. For both simulators, questionnaires were independently answered by the residents immediately after finishing the procedure. The questions were sub-divided into three main fidelity domains (environment, equipment, and psychological), as well as into questions pertaining to learning and the future use of these simulators.

Reliability of simulator specific questionnaires

The internal consistency was calculated to ensure that questions from each domain accurately measured the same construct. For the Sawbones simulator, each domain had a Cronbach's $\alpha > 0.70$, with a Cronbach's α of 0.89 when all three domains were combined (overall fidelity). The virtual simulator demonstrated a Cronbach's α of 0.60 for environment, 0.78 for equipment, and 0.83 for psychological domains. The overall fidelity domain of the virtual simulator was 0.88, similar to the Sawbones simulator. Upon review of the questions pertaining to the domains, one question was worded poorly for the virtual environment fidelity, and when removed, the internal consistency of the virtual environment was > 0.80. The results from the use of the questionnaires demonstrate that fidelity assessments were reliably measured for both these simulators.

Comparing fidelity of the Sawbones and Virtual simulators

The questionnaire findings were analyzed between groups (first simulator used and experience level) and between simulators. As expected, no differences were observed between first simulator used groups. Interestingly, junior residents scored the virtual environment with a significantly higher score than the senior residents, with a large effect size difference ($d = 1.11$) and therefore a meaningful difference. This is likely due to the senior residents having had more experience with actual surgical environments than the junior residents.

Effect size is important to report in understanding how education may influence learning, and it has been suggested for decades to be used in all behavioural sciences. Effect size helps define the meaningfulness of statistically significant results [105], and it has been recommended by the APA task force on statistical inference to be reported for all primary outcomes [106]. As described previously, an effect size difference of 0.2 is small, 0.5 is medium, and 0.8 is large. In order to facilitate visualization of these results, Cohen (1988) described the percentage of non-overlap between group scores and their correlation to effect size differences (*d*). At a *d* of 0.0, the group scores will overlap completely, in other words they have 0% non-overlap. At a *d* of 0.2, there is 14.7% non-overlap between groups. At $d = 0.5$ there is 33.0% non-overlap, and at $d = 0.8$ there is 47.4% non-overlap. This study found some very large effect size differences. The highest level stated by Cohen was $d = 2.0$, where there is 81.1% non-overlap of scores between groups. This indicates that in some of our results, the magnitude of the statistical difference was so large that over 80% of the group scores did not overlap, suggesting a very meaningful difference or change in learning outcomes between and within groups.

The fidelity of the virtual simulator was compared directly to the Sawbones simulator for each domain, as well as compared to a combination of all three domains (overall fidelity). For each of these comparisons, the participants assigned the Sawbones simulator a significantly higher level of fidelity than the virtual simulator, all with large effect size differences ($d = 1.38$ environment; $d = 1.71$ equipment; $d = 1.96$ psychological; $d = 2.31$ overall) indicating that these results are meaningful and indicate a large difference between simulators. The individual domains were also compared within each simulator. The Sawbones simulator environment and equipment domains were similarly rated (~81/100), and both were significantly higher than the psychological fidelity (72/100). The virtual simulator domains were each rated significantly different from each other, with the environment domain (68/100) > equipment domain (60/100) > psychological domain (50/100). Both simulators demonstrated that the psychological domain is the most difficult fidelity domain to re-create, while the environment is possibly the easiest.

Each question was also compared separately between simulators. The Sawbones simulator demonstrated significantly better scores than the virtual simulator on 60% of the environment domain questions, on 83% of equipment domain questions, and on 100% of psychological domain questions. This again reflects that the virtual environment was closest to real life (and the gold standard) and possibly easier to re-create, while the psychological fidelity domain requires much more work to be comparable with the gold standard.

It can be argued that high levels of fidelity are not important for all simulators, especially at junior training levels [107]. Virtual reality simulators, at this time, appear to be better for junior residents who are learning or practicing basic surgical skills. Higher fidelity simulators are ideal for more advanced surgical skills that require multiple tasks, and more experienced surgical trainees will likely benefit more from these simulators. Neither of these simulators received a high level of psychological fidelity, which may be acceptable at the training level. The procedure being performed on both simulators is a basic orthopaedic surgical skill, and participants felt that both of these simulators would be effective for the introduction of basic surgical skill (4.6/5 Sawbones, 4.2/5 virtual). However, the learner satisfaction questions demonstrated that participants felt the Sawbones simulator was more useful than the virtual simulator for practicing previously learned skills, for simulator-based examinations and for refresher skills, possibly due to its higher levels of overall fidelity.

When the participants were asked which surgical simulator they would use in their spare-time to practice surgical fixation of the ulna, the majority stated they would use the Sawbones instead of the virtual, as seen in other virtual simulator studies [57]. Most participants felt the virtual simulator still required further development prior to its use (4.6/5) whereas participants disagreed that Sawbones required further development (2.8/5). It was generally felt that both simulators provided a valuable experience and should be included in the residency-training program (4.1/5 Sawbones; 3.3/5 virtual), along with a surgical skills center (4.8/5), which are currently not available at the University of Calgary.

The fidelity of the Sawbones simulator is significantly higher than the virtual simulator in all domains, all with large effect size differences, signifying meaningful differences for future medical education initiatives. As a result, we can conclude that the newly created virtual simulator does not maintain a level of fidelity that is comparable with that of the current gold standard for surgical fixation of the ulna, for both junior and senior residents. The participants felt that further development of the virtual simulator would be beneficial prior to it being used for learning basic surgical skills. The participants also provided suggestions on how to improve this virtual model, which will aid in assessing the feasibility of developing a virtual surgical simulator (discussed below).

Question 2: Measurement of participant performance

Multiple measurement tools were developed for this study to assess participant performance and skill level, as well as the validity of the simulators. These include: 1) task-specific itemized checklist; 2) Global Rating Scale (GRS), 3) total errors, and 4) time to completion of the procedure. Before these tools can be used to evaluate participants and simulators, they must be proven reliable and valid.

Reliability of Measurement tools

Internal consistency

The internal consistency can be assessed using statistical analyses for newly constructed measurement tools. A Cronbach's α of 0.70 is the reported acceptable value identified a tool's reliability [84]. However, if a tool is to be used for an accrediting examination, a minimum Cronbach's α value of 0.80 or greater is usually sought. The internal consistency of both the newly created checklist and GRS were assessed for both simulators separately. The checklists for each simulator were both under the acceptable scale values, however, the GRS's were both above 0.70. In previous literature, the internal consistency of checklists alone ranged from $\alpha = 0.33 – 0.79$, while the GRS alone ranges from $\alpha = 0.66 – 0.98$ [4,72,77,80,81]. As with other studies, the internal consistency of the checklist appears to be lower than the GRS [72,78,81]. When both the checklist and GRS are combined, the findings for both simulators were above $\alpha = 0.80$. When the checklist and GRS are used in combination, these measurement tools have high internal consistency.

Inter-rater reliability

As this experiment used a modified version of a checklist and GRS, and was evaluating a new simulator, inter-rater reliability was assessed. A significant inter-rater reliability was noted for both simulators checklists with intraclass coefficient's (ICC) of >

0.70, but the GRS and total errors for both simulators were very low, ICC < 0.25. This may be due to both the inability of examiners to agree (poor inter-rater reliability), and to the variability in the way they score each participant. It has been noted in prior studies that low inter-rater reliability may be the result of decreased variability within one examiner's scores (i.e., always giving the median available score) [108]. Low inter-rater reliability may also be the result of inadequate examiner preparation or the lack of examiner experience [80]. This may be a concern in this study as examiners were only permitted ten minutes of training prior to their commencement. Examiner fatigue may also be a reason for low inter-rater reliability. The examiners were required to spend all day in the same room, evaluating participants on the same simulator, which can be mentally exhausting.

Studies consistently show inter-rater reliability for the checklist between 0.64 and 0.98, and for the GRS between 0.42 and 0.98 [4,71,75,80,85]. An accepted quality benchmark level of > 0.80 for inter-rater reliability [61] was met in this study when the checklist and GRS are combined, with an ICC of > 0.90. Previous studies have suggested that only one examiner needs to be used for OSATS type evaluations [4,71,75,109], and this data also suggests that for both of these simulators, only one examiner is needed when using both the checklist and GRS together.

In this study, neither examiner was completely blinded to participants' experience level, as they were all from the same orthopaedic surgery group. This likely did not affect the results as previous studies have shown that there are no significant differences between unblinded and blinded examiners, although there is a trend for unblinded examiners to give participants slightly higher scores [85].

Some studies suggest that the checklist is inferior to GRS [4,110]. A checklist simply asks if tasks were done correctly or not, whereas a GRS can evaluate how well the procedure was actually performed overall, since it is using a 5-point Likert scale. It has been suggested that the checklists are more useful for junior level trainees, and that the global rating scales are more useful for senior trainees [77]. Since the internal consistency is low for the checklists while being high for the GRS, and the results of the inter-rater reliability are opposite, one can infer that both the checklist and GRS should be used in combination for increased accuracy and reliability when assessing simulators such as these.

The reliability of using total errors was also assessed using inter-rater reliability. The inter-rater reliabilities of the total errors committed during the procedure for both simulators were very low (ICC Sawbones simulator < 0.0; ICC virtual simulator = 0.19). This suggests that total errors may not be a reliable measurement tool for assessing surgical fixation of the ulna. Total errors are used often as a measurement for determining the validity of new simulators [2,21,55,111,112], although a recent meta-analysis on laparoscopic virtual simulators has shown that there are no significant differences in performance errors noted for novices when comparing training on virtual laparoscopic simulators to training on gold standard simulators or no training at all [113]. If total errors are to be used for a measurement of a simulator's validity, caution should be exercised and a sufficient examiner-training period should be implemented to ensure that each examiner evaluates errors in the same fashion.

Three of the four measurement tools were evaluated for reliability. The checklist and GRS when used together demonstrate both high internal consistency and inter-rater reliability. The total errors score did not received a satisfactory intraclass coefficient, and thus cannot be judged as a reliable measure of the participants' performance. These tools were also evaluated in various ways to assess their validity.

Validity of Measurement tools

Criterion validity

One method to measure a tool's validity is to assess it with a comparable measure as part of a criterion validity test. This is the agreement between results of the newly created virtual simulator, and those of the gold standard (Sawbones). The Pearson's product-moment correlation coefficient can be used to determine the criterion validity of the measurement tools by assessing within simulator correlations between tools. The correlation coefficients between the checklist and GRS were significant at $r = 0.69$ and 0.57 for the Sawbones and virtual simulators respectively, which is lower than other studies determining the same correlations ($r = 0.81$-0.89) [4,71]. Significant correlations were also noted between the checklist and GRS to both total errors (virtual and Sawbones) and time to completion (Sawbones), as seen in Table 34. Similar to other studies, this study demonstrates a significant negative correlation between the checklist and GRS scores to both total errors and time to completion, ranging from $r = -0.48$ to - 0.81 [108].

The checklist and GRS were found to be reliable measures when used in combination, as they complement each other. The total errors (both simulators) and time to completion (Sawbones only) have demonstrated statistically significant criterion validity when correlating to the checklist and GRS measures. It may therefore be suggested that all four-measurement tools are reliable in assessing surgical fixation of the ulna for the Sawbones simulator, while the checklist, GRS and total error scores are reliable for the virtual simulator.

Validity of Simulators

Criterion validity

No statistically significant correlations between the measurement tools of the Sawbones and virtual simulators (between simulators analysis) were noted. This demonstrates that while each measurement tool is correlated within simulators, between simulator correlations are absent. This may indicate that the simulators are not measuring the same constructs accurately. When determining the validity of a simulator, more than one type of validity should be assessed, therefore construct validity was also evaluated.

Construct validity

Construct validity of simulators in surgical education can be described as the ability of a simulator to distinguish among differing experience levels. The construct validity of the Sawbones simulator for surgical fixation of an ulna has not been previously described. Using the four different objective measurements, the Sawbones simulator was able to accurately distinguish junior residents from senior residents. More statistically significant findings were noted when comparing junior and senior residents than when evaluated across PGY levels. This is likely due to insufficient numbers of participants within each year to accurately discriminate between years, and therefore comparing junior to senior residents may be more useful as seen in other studies, [4,82,114]. The newly developed virtual surgical simulator for surgical fixation of the ulna was able to distinguish junior residents from senior residents with the GRS and total errors. Senior residents obtained significantly higher scores on the GRS and committed significantly fewer errors during the procedure. When statistical significance was reached, large effect size differences (all between d = 0.98 and 1.98) were noted for both the Sawbones and virtual simulators, again indicating a meaningful difference in the participant's performance on the simulators when comparing experience levels.

Interestingly, the PGY-5 residents performed more poorly those more junior (PGY-3/4) in virtual GRS scores, virtual and Sawbones total errors and virtual time. The reasoning behind this is unclear but may be explained by negative transfer. Previous studies have shown prior experience in general surgery do not necessarily result in a positive transfer of skill when learning new tasks [115]. It has been shown that those participants that held no prior experience in microsurgery skills learned new skills more quickly, and that experience may have initially led to a negative transfer of skills for the previously experienced [115,116].

The measurement tools are reliable for both simulators, and may be used to accurately measure participant performance. The tools were also able to aid in determining quality of surgical skill (especially the questions on the GRS), as can be noted by the construct validity of the simulators. The virtual simulator for surgical fixation of the ulna was, however, unable to achieve criterion validity when correlating it to the gold standard Sawbones simulator, and it was only able to demonstrate construct validity for two of the four measurement tools. As a result, the tools are reliable in measuring surgical fixation of the ulna, but the virtual simulator may not be accurately replicating the same construct as the Sawbones simulator.

Question 3: Learners' perception

Learning in surgical education can be divided into three general domains: knowledge, skill and attitude (or in the case of this present study, level of comfort). These three domains were evaluated by the participants before and after the procedures, to determine if any self-assessed improvement had occurred. The overall results of the post-procedure assessment were found to be significantly higher than the pre-procedure assessment, although with a small effect size difference ($d = 0.20$). Since this effect size was small, each post-graduate year (PGY) was evaluated separately. PGY-1 residents were the only group to show a statistically significant improvement on the post-procedure questionnaire, with an overall large effect size difference ($d = 1.30$).

The three domains of learning were also analyzed separately. For the skill and comfort domains, significant improvements in scores were noted for the entire cohort, both with small effect size differences ($d = 0.18$ and $d = 0.22$ respectively). When evaluating each domain among experience levels (PGY) separately, PGY-1's were once again the only experience level where significant improvements were noted. Both the skill and comfort domains showed significant improvements with large effect size differences ($d = 1.20$ and $d = 1.55$ respectively), indicating there was a meaningfully large improvement in performance between pre and post-questionnaire scores for the PGY-1 residents.

Construct validity

The construct validity of the pre and post-procedure questionnaires were assessed based on the residents' level of experience as a function of their year in the orthopaedic residency program. The pre-procedure questionnaire was best able to differentiate between experience levels (PGY-1 scores < PGY-2, 3 < PGY-4, 5). Both the pre and post-procedure questionnaires were able to distinguish junior residents (PGY-1, 2) from senior residents (PGY-3, 4,5) showing a very large effect size difference (pre-questionnaire, $d = 2.41$; post-questionnaire, $d = 2.56$), indicating that there was a meaningful difference in scores between the two experience levels. When each domain was evaluated separately, junior residents again demonstrated a significantly lower score than senior residents, each with large effect size differences (pre-knowledge, d = 2.40; post-knowledge, d = 2.19; pre-skills, d = 2.32; post-skill, 2.17; pre-comfort, 2.28; post-comfort, 2.68).

The pre-procedure questionnaire also asked about participant demographics and video game experience. The amount of time spent playing video games in the past and present were correlated to all tools, including fidelity questionnaires. There appeared to be no influence of video game use and performance, unlike other studies [34].

Question 4: Feasibility

Participant feedback

Participant feedback is essential when creating a new simulator and determining the feasibility of its development. Short answer questions were created and included in the simulator specific post-procedure questionnaires. Common advantages among the simulators were that they are both ideal for reviewing and learning basic steps of the procedure, and for familiarizing oneself with the equipment associated with the procedure. The visualization of both simulators tools were also noted as a benefit. Common weaknesses of the simulators are that neither are realistic enough in comparison to operating on a live patient, both are “too easy”, and neither evoked the stress or anxiety normally felt by the resident in the operating room. In addition, there was no fracture to reduce on either simulator. Noted weaknesses of the virtual simulator were incorrect tactile feedback for drilling/tapping/screwing, lack of sound, lack of using two hands, and no stereoscopic vision. Similar simulator frustrations were that the simulated bones were unlike real bone for drilling, and had no soft tissue envelope for dissection. The virtual simulator was frustrating for many participants due to the visualization of tools being used (i.e., the drill did not always enter the correct spot on bone or plate), and a learning curve was associated with tool and computer usage. Importantly, changes to the virtual simulator were addressed. Improving the haptic feedback, using real tool handles, and adding sound were among the most common comments from the residents, along with performing the actual reduction, using two haptics devices, correlating visual and haptics more efficiently, and adding stereoscopic vision. As important as sound is thought to be for these participants, it may not be detrimental to simulators for novices practicing basic skills such as this one. It has been suggested that novice trainees do not require drill noises to guide drilling, whereas it becomes more beneficial for intermediate trainees and experts [117].

Future benefits of both types of simulators included their use for learning basic psychomotor skills in a laboratory. Both are safe, can be used in a non-threatening environment, and the virtual simulator is much cleaner and easier to set up and take down than a Sawbones simulator. Benefits of virtual simulators are that they can eventually produce complex fractures or procedures, and create scenarios that may be repeated as often as able, while benefits of Sawbones simulators include shorter associated learning curves, and learning which instruments to use and how.

Cost analysis

Many studies have demonstrated that virtual simulators are as effective as standard simulators for developing and evaluating surgical skills, although this virtual surgical simulator still requires more work to reach the same level as the current gold standard.

When a new simulator is comparable to the gold standard, and if surgical trainees consider it useful and are willing to use it for learning and practicing procedures (as with this virtual simulator), an important factor in choosing one simulator over another is cost. A cost analysis was therefore performed to provide further comparisons between these simulators.

The largest costs for a virtual surgical simulator are the start-up costs, as previously thought [118]. This simulator, using a sophisticated six degrees of freedom haptics device, cost approximately $85,000.00 to develop. This included costs of personnel time for the development and maintenance of the simulator. In comparison, purchasing a Sawbones simulator for surgical fixation of the ulna would cost approximately $3,445.00. This does not include a Sawbones (ulna) for each resident (add approximately $780.00). This low cost is due to the prior development of the Sawbones modeled ulna and it's equipment.

Over a 5-year period the costs of a virtual simulator does not increase greatly. The only requirements are upgrades to the haptics device and periodic maintenance, bringing a 5-year cost total to approximately $105,000.00. The Sawbones simulator costs will increase more rapidly, as tools will need to be replaced, and more Sawbones ulna models will need to be purchased and shipped. A 5-year cost for a Sawbones simulator for surgical fixation of the ulna, with each resident only practicing on ten Sawbones per year, would be approximately $55,150.00. Continuing with this trend, by ten years the cost gap closes (virtual $125,000.00 vs. Sawbones $110,300.00), and by 15 years the virtual simulator will cost approximately $145,000.00, while the Sawbones simulator will approach $165,450.00. If residents were to use 20 Sawbones per year instead of 10, this would increase the 5-year cost to $94,150.00, and the costs would be greater than the virtual model by year 10.

Another method of practicing surgical fixation of the ulna is to rent the equipment. For a 4-hour, 4 person course, the costs would be approximately $2000.00. For each resident to attend twice in a year, the costs would be approximately $24,000.00. In five years, this cost would be $120,000.00, well over that of the virtual simulator costs.

Sawbones are often used for fracture fixation courses for residents and surgeons alike. These courses usually involve around 20 pairs of participants practicing the surgical fixation all at one time, each with their own simulator station, costing approximately $3,500.00 per pair (approximately $35,000.00). In order to perform a similar course using a virtual haptics model, you would need at least 10 haptics devices and computers, each costing approximately $65,000.00 (approximately $650,000.00). Buying these haptics devices in bulk may result in some discount, but at this time, a course with virtual simulators would be much more expensive than a Sawbones course.

The long term monetary savings with virtual simulators is noted, with the added bonus of being allowed to practice more than ten times in a year. Of course, the advantages of repeated practice time without added cost, and with the possibility for the simulator to provide immediate objective feedback and in the future incorporate surgical difficulties or more complicated fracture patterns, may provide the participant with more unique reasons to use virtual simulators in the future.

Exit interviews

Important to the assessment of the feasibility of creating a high fidelity virtual surgical simulator is the insight of the developers. This may aid in the development of other simulators by other groups. Both of the developers for this virtual simulator were engineering students: one a final year undergraduate and the other a 1st year master’s student. Neither had previous experience in creating virtual simulators, although they had access to supervisors with prior experience. A total of approximately 776 hours were spent in learning software, creating the simulator, and performing minor changes or upgrades. Both developers were satisfied with the simulator prior to its use in the study, although they felt that with prior knowledge of virtual environments and its computer code, a more accurate simulator might have been created in a similar timeline. The easiest aspect of the development of the simulator was defining and dealing with graphical objects and visualizations, while the hardest tasks were pertaining to the rendering of force feedback sensations.

Limitations

The major limitations to this study were funding for more advanced equipment and the use of senior simulator developers. These have an impact on achieving better fidelity of the simulator. The participating engineering supervisor and laboratory have been involved in previous surgical simulations in thoracic surgery [114 119]. They already had all the necessary basic equipment and software to develop this simulator in their laboratory. Additional haptics devices and newer devices that could be purchased to increase the fidelity of the model were not available for this study. Another limitation to this study was the time given to complete the simulator. The goal of this study was to assess if a high fidelity surgical simulator with haptics could be developed in one year. If more time was provided, perhaps a higher quality surgical simulator may have been developed for use in this study. Using a simulator for a first time is also a limiting factor. Many simulators have learning curves associated with them. Due to costs and time, it was only feasible for the participants to practice this procedure on the virtual simulator for ten minutes.

The virtual checklist also had some limitations, as a significantly negative skewness was noted, indicating that the majority of the participants attained high scores. This may have been due to the inability to distinguish if a task was completed correctly or incorrectly, or because the procedure/simulator was too straightforward. During the virtual simulation, it is possible that examiners were unable to accurately assess the quality of the tasks and whether or not they were performed correctly. This can be, however, more accurately assessed on the Sawbones simulator.

All the above mentioned limitations to the study have influenced the ability to create a high fidelity surgical simulator for surgical fixation of the ulna.

Conclusion

Benefits of non-medical simulators have been known for decades. The military have determined that a soldier has a 95% chance of returning home alive after they have completed 10 missions; therefore their first 10 missions are best initially completed in a simulator [24]. This may hold true in the medical community as well, where initial procedures can be safely practiced and learned on simulators prior to being performed on live human beings. Surgical simulators are being developed to aid surgical trainees to repeatedly and safely practice psychomotor skills in a non-threatening environment during the steepest portion of their learning or performance curve. The feasibility and fidelity of a newly created virtual simulator for surgical fixation of an ulna was assessed in this study.

Multiple measurement tools were developed to help in this assessment, including task-specific checklists, global rating scales, and post-procedure questionnaires specifically designed to assess the simulators fidelity. A questionnaire was also developed to assess the participants' knowledge, skill and comfort of surgical fixation of an ulna prior to and after the procedure to determine if any improvement occurred in these domains. Cost analysis and an exit interview were also conducted to aid in determining the virtual simulators feasibility of being developed within a year.

This is the first study to assess the fidelity of a Sawbones simulator, using questionnaires that were specifically created to assess three separate fidelity domains of simulators. The baseline fidelity of the current gold standard for learning and practicing surgical fixation of the ulna was used to compare the fidelity of the newly developed virtual ulna fracture fixation simulator. The questionnaire results demonstrated that the overall fidelity of the virtual simulator (60/100) was significantly lower than the Sawbones simulator (79/100), with significant differences noted for each separate fidelity domain. The level of fidelity in the new virtual simulator does not meet the same standards as the Sawbones simulator.

Construct validity of the simulators was confirmed, including the Sawbones simulator, which prior to this study had yet to be verified. When using the GRS and total error scores with the virtual simulator, significant differences were noted between junior and senior residents, and from the large effect size differences found are therefore meaningful from an educational perspective. As discussed in the limitations, the checklist was noted to have a significant skewness which may be due to the examiner's inability to correctly assess if a task was completed correctly, leading the examiner to mark the task as done, rather than completed properly. In addition, the time to completion of the virtual procedure did not achieve construct validity (i.e., no performance differences were found between junior and senior residents), although this may not be entirely to do with the simulator itself. Superior scores on time to completion of the procedure may be related to one's previous ability to use the instruments of the virtual and Sawbones simulators. An example of this is the PGY-5's who have used Sawbones multiple times prior to this procedure. They had the fastest times to completion on the Sawbones simulator as they are at the peak of their performance curve. Meanwhile, having used the virtual simulator for the first time, they shared some of the longest times to completing the procedure on this simulator (where they are still at the steep portion of their learning or performance curve). Time to completion of a procedure alone may not be the best indicator of the construct validity of a simulator. A surgeon may perform a procedure quickly, but it may be executed unsafely and with poor results. Outcomes may actually improve when more time is taken to perform a procedure. Time to completion is used frequently to assess if virtual simulators are valid and reliable tools for surgical education [1,2,9,17,18,31,57]. It can, however, be argued that time to completion is an important outcome, as participants who have the knowledge, skill and comfort levels needed to correctly perform a procedure will eventually become more proficient and ultimately be able to perform the task faster. Therefore, it is not incorrect to utilize time as a measurement tool, but it should not be used in isolation for the validation of a simulator. It should be correlated with additional measurements such as quality and patient safety outcomes.

When combined, both the internal consistency and inter-rater reliability of the checklist and GRS was high as in other studies, while the inter-rater reliability of a total error score was low. Even though total errors correlated highly with both the checklist and GRS, the low inter-rater reliability may suggest that total error scores are more difficult to assess, and caution should be exercised when using it to evaluate simulators. Therefore, it can be suggested that the checklist and GRS when used collectively have been shown to be reliable and are beneficial for the assessment for surgical fixation of the ulna. Total error scores and time to completion are valuable as well, but should not be used in isolation to assess a new simulator.

When comparing the average scores of all measured data, the virtual simulator scores ranked higher than the Sawbones simulator in all cases except time. With the measurement tools showing non-significant correlations between simulators, and the fidelity of the virtual simulator being significantly lower than the Sawbones simulator, these increased scores for the virtual simulator are likely due to the fact that the newly created virtual simulator does not meet the same standards as the current gold standard for surgical fixation of the ulna. The newly created measurement tools developed for this study were shown to reliably assess surgical fixation of the ulna for both simulators. In addition, the virtual simulator does demonstrate construct validity with the ability to distinguish between experience levels for both the GRS and total error scores. However, increasing the overall fidelity of the simulator is required prior to it being used as a training tool for the surgical fixation of the ulna.

This simulator was developed with the intended use for surgical education. The developed questionnaire assessed the participants' knowledge, skill and comfort with surgical fixation of the ulna; showing overall improvement after the residents had completed the procedures on both simulators. This short questionnaire may be valuable in the future when assessing how training on surgical simulators affects the participants' self-perceived knowledge, skills and comforts with other procedures.

As stated in other studies, virtual simulators may save money with long-term usage. The start up costs for a virtual simulator are much higher than buying all the materials required for a Sawbones simulator, but over 10-15 years the costs of replacing tools and buying Sawbones will surpass the costs of upgrading and maintaining a virtual simulator.

Once this virtual simulator has been upgraded and its fidelity improved, it may be as beneficial to surgical trainees as a Sawbones simulator, and may be a cheaper method of training over time.

We have determined that it is not feasible at this time to develop a high fidelity virtual simulator for surgical fixation of the ulna in one year. In order to accomplish this, previous knowledge and experience with developing virtual simulators and their software is essential. The tools to assess surgical fixation of an ulna for both Sawbones and virtual simulators have been developed however, and have been proven reliable, and may be used for future comparisons.

Future directions

At this time, the cohort of orthopaedic surgery residents that participated in this study preferred to use the gold standard Sawbones simulator for surgical fixation of the ulna; mostly due to the realism of the tools being used, the ability to use both hands simultaneously, and the benefits of hearing sounds produced as the fixation of the ulna is being completed. This has also been noted with general surgery residents who favor video box trainers to virtual laparoscopic simulators [120]. With more experience in developing these virtual simulators, increased funding and the advent of new software/hardware, the virtual simulators may one day match our current gold standards.

The ultimate goal is not to replace the current simulators or methods of teaching surgical skills, but to use virtual simulators as an additional resource for training. Not all trainees develop skills in the same fashion, so multiple methods should be available. It has been suggested that distributed learning, and not mass learning, is better suited with the retention of knowledge [100]. A virtual simulator, which is available at all times, is better suited for this type of learning. With ease of availability, decreased set up and take down time, a surgical trainee can have frequent, short practice sessions at anytime, rather than only a few times a year for extended hours. The additional advantages of virtual trainers are many, including the promising role of a simulators ability to objectively evaluate surgical performance and competence. This training method may also help to evaluate if surgical trainees have reached a minimum level of competency to allow them to perform procedures in the operating room [107].

The medical community is a group that hesitates to change without high quality evidence based medicine. A high fidelity, validated simulator that can accurately evaluate a surgical trainee and also demonstrate transfer of skills to a real life operation is the ultimate goal for surgical education. There are a multitude of virtual surgical simulators available, however, many of these need to be properly evaluated to determine what upgrades are required to attain this goal. With the rise of competent new surgeons who grew up in a world of advanced video games and computers, the advent of high fidelity simulators may become an ideal method for learning and practicing surgical skills to develop and enhance proficiency outside of the operating room.

REFERENCES

1. Gomoll AH, O'Toole RV, Czarnecki J, Warner JJP. Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. *The American Journal of Sports Medicine 2007;35-6:883.*

2. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British journal of surgery 2004;91-2:146-50.*

3. Berg D, Raugi G, Gladstone H, Berkley J, Weghorst S, Ganter M, Turkiyyah G. Virtual reality simulators for dermatologic surgery: measuring their validity as a teaching tool. *Dermatologic Surgery 2001;27-4:370-4.*

4. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. *British journal of surgery 1997;84-2:273-8.*

5. Müller WK, Ziegler R, Bauer A, Soldner EH. Virtual reality in surgical arthroscopic training. *Journal of Image Guided Surgery 1995;1-5:288-94.*

6. Cannon WD, Eckhoff DG, Hunter RE, Sweeney HJ. Report of a group developing a virtual reality simulator for arthroscopic surgery of the knee joint. *Clinical Orthopaedics and Related Research 2006;442:21.*

7. Lemole Jr GM, Banerjee PP, Luciano C, Neckrysh S, Charbel FT. Virtual reality in neurosurgical education: part-task ventriculostomy simulation with dynamic visual and haptic feedback. *Neurosurgery 2007;61-1:142.*

8. Wang P, Becker AA, Jones IA, Glover AT, Benford SD, Greenhalgh CM, Vloeberghs M. A virtual reality surgery simulation of cutting and retraction in neurosurgery with force-feedback. *Computer methods and programs in biomedicine 2006;84-1:11-8.*

9. Hogle NJ, Briggs WM, Fowler DL. Documenting a learning curve and test-retest reliability of two tasks on a virtual reality training simulator in laparoscopic surgery. *Journal of Surgical Education 2007;64-6:424-30.*

10. Morris D, Sewell C, Barbagli F, Salisbury K, Blevins NH, Girod S. Visuohaptic simulation of bone surgery for training and evaluation. *IEEE Computer Graphics and Applications 2006:48-57.*

11. Zhong Y, Shirinzadeh B, Alici G, Smith J. Soft tissue modelling through autowaves for surgery simulation. *Medical and Biological Engineering and Computing 2006;44-9:805-21.*

12. Aggarwal R, Tully A, Grantcharov T, Larsen CR, Miskry T, Farthing A, Darzi A. Virtual reality simulation training can improve technical skills during laparoscopic salpingectomy for ectopic pregnancy. *BJOG(Oxford. Print) 2006;113-12:1382-7.*

13. Gorman PJ, Meier AH, Krummel TM. Simulation and Virtual Reality in Surgical Education Real or Unreal? *Archives of Surgery 1999;134-11:1203-8.*

14. Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. *Studies in health technology and informatics 1998;50:124.*

15. Sourin A, Sourina O, Sen HT. Virtual orthopedic surgery training. *IEEE Computer Graphics and Applications 2000;20-3:6-9.*

16. Mabrey JD, Gillogly SD, Kasser JR, Sweeney HJ, Zarins B, Mevis H, Garrett Jr WE, Poss R, Cannon WD. Virtual reality simulation of arthroscopy of the knee. *Arthroscopy: the journal of arthroscopic & related surgery: official publication of the Arthroscopy Association of North America and the International Arthroscopy Association 2002;18-6:E28.*

17. McCarthy AD, Moody L, Waterworth AR, Bickerstaff DR. Passive Haptics in a Knee Arthroscopy Simulator: Is It Valid for Core Skills Training? *Clinical Orthopaedics & Related Research 2006;442:13.*

18. Gomoll AH, Pappas G, Forsythe B, Warner JJP. Individual skill progression on a virtual reality simulator for shoulder arthroscopy: a 3-year follow-up study. *The American Journal of Sports Medicine 2008;36-6:1139.*

19. Pedowitz RA, Esch J, Snyder S. Evaluation of a virtual reality simulator for arthroscopy skills development. *Arthroscopy 2002;18-6:E29.*

20. Srivastava S, Youngblood PL, Rawn C, Hariri S, Heinrichs WL, Ladd AL. Initial evaluation of a shoulder arthroscopy simulator: establishing construct validity. *Journal of Shoulder and Elbow Surgery 2004;13-2:196-205.*

21. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of Surgery 2002;236-4:458.*

22. Strachan IW. Technology Leaps All Around Propel Advances in Simulators-Simulation and Training Industry: Changes in the Corporate Landscape. *National Defense 2000:86-93.*

23. Krummel TM. Surgical simulation and virtual reality: the coming revolution. *Annals of Surgery 1998;228-5:635.*

24. Rosen KR. The history of medical simulation. *Journal of Critical Care 2008;23-2:157.*

25. Explorations in Learning & Instruction: The Theory Into Practice Database *Experiential Learning (C. Rogers).* [updated 2009; cited 2008 December 8]; Available from: http://tip.psychology.org/rogers.html.

26. Bloom BS. Taxonomy of Educational Objectives. Handbook I: Cognitive Domain. New York, David McKay Co. *Inc., New York 1956.*

27. Krathwohl DR, Bloom BS, Masia BB. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain, David McKay Company. *Inc., New York 1964.*

28. Simpson E. The classification of educational objectives: psychomotor domain. *Illinois Journal of Home Economics 1966;10-4:110-44.*

29. Aggarwal R, Ward J, Balasundaram I, Sains P, Athanasiou T, Darzi A. Proving the Effectiveness of Virtual Reality Simulation for Training in Laparoscopic Surgery. *Annals of Surgery 2007;246-5:771.*

30. Verdaasdonk EGG, Stassen LPS, Monteny LJ, Dankelman J. Validation of a new basic virtual reality simulator for training of basic endoscopic skills. *Surgical endoscopy 2006;20-3:511-8.*

31. Aggarwal R, Hance J, Undre S, Ratnasothy J, Moorthy K, Chang A, Darzi A. Training junior operative residents in laparoscopic suturing skills is feasible and efficacious. *Surgery 2006;139-6:729-34.*

32. Wanzel KR, Ward M, Reznick RK. Teaching the surgical craft: from selection to certification. *Current problems in surgery 2002;39-6:573-660.*

33. Esen H. A control algorithm and preliminary user studies for a bone drilling medical training system. *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003.* 2003:153- 8.

34. Rosser Jr JC, Lynch PJ, Cuddihy L, Gentile DA, Klonsky J, Merrell R. The impact of video games on training surgeons in the 21st century. *Archives of Surgery 2007;142-2:181.*

35. Ström P, Kjellin A, Hedman L, Johnson E, Wredmark T, Felländer-Tsai L. Validation and learning in the Procedicus KSA virtual reality surgical simulator. *Surgical endoscopy 2003;17-2:227-31.*

36. Maran NJ, Glavin RJ. Low-to high-fidelity simulation-a continuum of medical education? *Medical Education 2003;37-s1:22-8.*

37. Alexander AL, Brunyé T, Sidman J, Weil SA, Aptima I, Woburn MA. From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *The Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2005.*

38. Srinivasan MA, Basdogan C. Haptics in virtual environments: Taxonomy, research status, and challenges. *Computers & Graphics 1997;21-4:393-404.*

39. Meehan M, Insko B, Whitton M, Brooks Jr FP. Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics (TOG) 2002;21-3:645-52.*

40. Chou B, Handa VL. Simulators and virtual reality in surgical education. *Obstetrics and gynecology clinics of North America 2006;33-2:283-96.*

41. Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, Jarvi KA. The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. *Annals of Surgery 2004;240-2:374.*

42. Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, Hutchison C. Assessment of technical skills transfer from the bench training model to the human model. *American journal of surgery 1999;177-2:167.*

43. Bradley P. The history of simulation in medical education and possible future directions. *Medical Education 2006;40-3:254-62.*

44. Cooper JB, Taqueti VR. A brief history of the development of mannequin simulators for clinical education and training. *British Medical Journal 2008;84-997:563.*

45. Wallace P, Patient S. Following the threads of an innovation: the history of standardized patients in medical education. *Caduceus 1997;13-2:5-28.*

46. Ackerman MJ. The visible human project. *Proceedings of the IEEE 1998;86-3:504-11.*

47. Schwid HA. Components of an effective medical simulation software solution. *Simulation & Gaming 2001;32-2:240.*

48. Arne R, Stale F, Ragna K, Petter L. PatSim-simulator for practising anaesthesia and intensive care. *International Journal of Clinical Monitoring and Computing 1995;13:147-52.*

49. Gilbart MK, Hutchison CR, Cusimano MD, Regehr G. A computer-based trauma simulator for teaching trauma management skills. *American journal of surgery 2000;179-3:223.*

50. Satava MR. Surgical education and surgical simulation. *World journal of surgery 2001;25-11:1484-9.*

51. Bridges M, Diamond DL. The financial impact of teaching surgical residents in the operating room. *The American Journal of Surgery 1999;177-1:28-32.*

52. Delp SL, Loan JP, Hoy MG, Zajac FE, Topp EL, Rosen JM. An interactive graphics-based model of the lower extremity to study orthopaedic surgical procedures. *IEEE transactions on bio-medical engineering 1990;37-8:757.*

53. Satava RM. Virtual reality surgical simulator: The first steps. *Surgical endoscopy 1993;7-3:203-5.*

54. Kothari SN, Kaplan BJ, DeMaria EJ, Broderick TJ, Merrell RC. Training in laparoscopic suturing skills using a new computer-based virtual reality simulator (MIST-VR) provides results comparable to those with an established pelvic trainer system. *Journal of Laparoendoscopic & Advanced Surgical Techniques 2002;12-3:167-73.*

55. Munz Y, Kumar BD, Moorthy K, Bann S, Darzi A. Laparoscopic virtual reality and box trainers: is one superior to the other? *Surgical endoscopy 2004;18-3:485-94.*

56. Torkington J, Smith SGT, Rees BI, Darzi A. Skill transfer from virtual reality to a real laparoscopic task. *Surgical endoscopy 2001;15-10:1076-9.*

57. Hamilton EC, Scott DJ, Fleming JB, Rege RV, Laycock R, Bergen PC, Tesfay ST, Jones DB. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surgical endoscopy 2002;16-3:406-11.*

58. Schijven MP, Jakimowicz JJ. Validation of virtual reality simulators: Key to the successful integration of a novel teaching technology into minimal access surgery. *Minimally Invasive Therapy and Allied Technologies 2005;14-4:244-6.*

59. Gay LR, Mills GE, Airasian P. *Educational Research: Competencies for analysis and application.* 9 ed. Upper Saddle RIver, New Jersey: Davis, K. M., 2009:648.

60. Lamata P, Gómez EJ, Sánchez-Margallo FM, Lamata F, del Pozo F, Usón J. Tissue consistency perception in laparoscopy to define the level of fidelity in virtual reality simulation. *Surgical endoscopy 2006;20-9:1368-75.*

61. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surgical endoscopy 2003;17-10:1525.*

62. Ferlitsch A, Glauninger P, Gupper A, Schillinger M, Haefner M, Gangl A, Schoefl R. Evaluation of a virtual endoscopy simulator for training in gastrointestinal endoscopy. *Endoscopy 2002;34-9:698.*

63. Kim HK, Rattner DW, Srinivasan MA. Virtual-reality-based laparoscopic surgical training: the role of simulation fidelity in haptic feedback. *Computer aided surgery: official journal of the International Society for Computer Aided Surgery 2004;9-5:227.*

64. Lendvay TS, Casale P, Sweet R, Peters C. Initial validation of a virtual-reality robotic simulator. *Journal of Robotic Surgery 2008;2-3:145-9.*

65. Mackay D, Wood L, Rangan A. The treatment of isolated ulnar fractures in adults: a systematic review. *Injury 2000;31-8:565-70.*

66. O'Toole RV, Playter RR, Krummel TM, Blank WC, Cornelius NH, Roberts WR, Bell WJ, Raibert M. Measuring and developing suturing technique with a virtual reality surgical simulator. *Journal of the American College of Surgeons 1999;189-1:114.*

67. Cristofolini L, Viceconti M. Mechanical validation of whole bone composite tibia models. *Journal of biomechanics 2000;33-3:279-88.*

68. Willis AA, Kutsumi K, Zobitz ME, Cooney Iii WP. Internal fixation of dorsally displaced fractures of the distal part of the radius. A biomechanical analysis of volar plate fracture stability. *The Journal of Bone and Joint Surgery 2006;88-11:2411.*

69. Zdero R, Rose S, Schemitsch EH, Papini M. Cortical screw pullout strength and effective shear stress in synthetic third generation composite femurs. *Journal of biomechanical engineering 2007;129:289.*

70. Reznick RK. Teaching and testing technical skills. *The American Journal of Surgery 1993;165-3:358-61.*

71. Winckel CP, Reznick RK, Cohen R, Taylor B. Reliability and construct validity of a structured technical skills assessment form. *American journal of surgery 1994;167-4:423.*

72. Anastakis DJ, Wanzel KR, Brown MH, Herold McIlroy J, Hamstra SJ, Ali J, Hutchison CR, Murnaghan J, Reznick RK, Regehr G. Evaluating the effectiveness of a 2-year curriculum in a surgical skills center. *The American Journal of Surgery 2003;185-4:378-85.*

73. Lentz GM, Mandel LS, Goff BA. A six-year study of surgical teaching and skills evaluation for obstetric/gynecologic residents in porcine and inanimate surgical models. *American journal of obstetrics and gynecology 2005;193-6:2056-61.*

74. Datta V, Chang A, Mackay S, Darzi A. The relationship between motion analysis and surgical technical assessments. *American journal of surgery 2002;184-1:70.*

75. Goff BA, Lentz GM, Lee D, Houmard B, Mandel LS. Development of an objective structured assessment of technical skills for obstetric and gynecology residents. *Obstetrics and gynecology(New York. 1953) 2000;96-1:146-50.*

76. Nielsen PE, Foglia LM, Mandel LS, Chow GE. Objective structured assessment of technical skills for episiotomy repair. *American journal of obstetrics and gynecology 2003;189-5:1257-60.*

77. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic medicine: journal of the Association of American Medical Colleges 1998;73-9:993.*

78. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative" bench station" examination. *American journal of surgery 1997;173-3:226.*

79. Swift SE, Carter JF. Institution and validation of an observed structured assessment of technical skills (OSATS) for obstetrics and gynecology residents and faculty. *American Journal of Obstetrics & Gynecology 2006;195-2:617.*

80. VanBlaricom AL, Goff BA, Chinn M, Icasiano MM, Nielsen P, Mandel L. A new curriculum for hysteroscopy training as demonstrated by an objective structured assessment of technical skills (OSATS). *American journal of obstetrics and gynecology 2005;193-5:1856-65.*

81. Ault G, Reznick R, MacRae H, Leadbetter W, DaRosa D, Joehl R, Peters J, Regehr G. Exporting a technical skills evaluation technology to other sites. *American journal of surgery 2001;182-3:254.*

82. Brydges R, Classen R, Larmer J, Xeroulis G, Dubrowski A. Computer-assisted assessment of one-handed knot tying skills performed within various contexts: a construct validity study. *The American Journal of Surgery 2006;192-1:109-13.*

83. Bodle JF, Kaufmann SJ, Bisson D, Nathanson B, Binney DM. Value and face validity of objective structured assessment of technical skills (OSATS) for work based assessment of surgical skills in obstetrics and gynaecology. *Med Teach 2008;30-2:212-6.*

84. Nunally JC, Bernstein IH. Psychometric theory. New York: McGraw-Hill, 1978.

85. Goff B, Mandel L, Lentz G, VanBlaricom A, Oelschlager AMA, Lee D, Galakatos A, Davies M, Nielsen P. Assessment of resident surgical skills: is testing feasible? *American journal of obstetrics and gynecology 2005;192-4:1331-8.*

86. Kohls-Gatzoulis JA, Regehr G, Hutchison C. Teaching cognitive skills improves learning in surgical skills courses: a blinded, prospective, randomized study. *Can J Surg 2004;47-4:277-83.*

87. Dieckmann P, Manser T, Wehner T. Presence and High Fidelity Patient Simulators in Anaesthesiology. Influences Derived From Interviews and Questionnaires. *6th International Workshop on Presence,.* Aalborg, 2003:6-8.

88. Olympio MA, Whelan R, Ford RPA, Saunders ICM. Failure of simulation training to change residents' management of oesophageal intubation. *British journal of anaesthesia 2003;91-3:312-8.*

89. Savoldelli G, Ed M, Naik V, Joo H, Houston P, Graham M, Yee B, Hamstra SJ. Evaluation of Patient Simulator Performance as an Adjunct to the Oral Examination for Senior Anesthesia Residents. *Anesthesiology 2006;104-3:475.*

90. Tsai MD, Hsieh MS, Jou SB. Virtual reality orthopedic surgery simulator. *Computers in biology and medicine 2001;31-5:333.*

91. Via DK, Kyle RR, Trask JD, Shields CH, Mongan PD. Using high-fidelity patient simulation and an advanced distance education network to teach pharmacology to second-year medical students. *Journal of clinical anesthesia 2004;16-2:144-51.*

92. Wright SW, Lindsell CJ, Hinckley WR, Williams A, Holland C, Lewis CH, Heimburger G. High fidelity medical simulation in the difficult environment of a helicopter: feasibility, self-efficacy and cost. *BMC Medical Education 2006;6-1:49.*

93. Lighthall GK, Barr J, Howard SK, Gellar E, Sowb Y, Bertacini E, Gaba D. Use of a fully simulated intensive care unit environment for critical event management training for internal medicine residents*. *Critical Care Medicine 2003;31-10:2437.*

94. Reich O, Noll M, Gratzke C, Bachmann A, Waidelich R, Seitz M, Schlenker B, Baumgartner R, Hofstetter A, Stief CG. High-level virtual reality simulator for endourologic procedures of lower urinary tract. *Urology 2006;67-6:1144-8.*

95. Lok B, Naik S, Whitton M, Brooks FP. Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence: Teleoperators & Virtual Environments 2003;12-6:615-28.*

96. Usoh M, Catena E, Arman S, Slater M. Using Presence Questionnaires in Reality. *Presence: Teleoperators and Virtual Environments 2000;9-5:497-503.*

97. Witmer BG, Singer MJ. Measuring presence in virtual environments: A presence questionnaire. *Presence 1998;7-3:225-40*.

98. Taekman JM, Hobbs G, Barber L, Phillips-Bute BG, Wright MC, Newman MF, Stafford-Smith M. Preliminary report on the use of high-fidelity simulation in the training of study coordinators conducting a clinical research protocol. *Anesthesia and analgesia 2004;99-2:521-7.*

99. Sawbones. SAWBONES worldwide, Division of Pacific Research Laboratories Inc. Vol. 2009: Division of Pacific Research Laboratories Inc., 2009. [updated 2009; cited 2009 May 25]; Available from: http://www.sawbones.com/default.aspx.

100. Mackay S, Morgan P, Datta V, Chang A, Darzi A. Practice distribution in procedural skills training. *Surgical endoscopy 2002;16-6:957-61.*

101. Brant R. Inference for a Mean: Comparing a Mean to a Known Value. [text/html]; [Sept 30, 2008; cited 2008 June 30]; Sample Size calculator]. Available from: http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html.

102. Daniels M. Psychic Science Random Number Generator and Checker. Vol. 2009. [May 21, 2009; cited 2009 April 1]; Available from: http://www.psychicscience.org/random.aspx.

103. Cohen J. *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.

104. Tabachnick BG, Fidell LS. Using multivariate statistics 3rd ed. 1996 Harper Collins College Publishers. *New York, NY.*

105. Olejnik S, Algina J. Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology 2000;25-3:241-86.*

106. Wilkinson L. Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist 1999;54:594-604.*

107. Reznick RK, MacRae H. Teaching surgical skills--changes in the wind. *The New England journal of medicine 2006;355-25:2664.*

108. Szalay D, MacRae H, Regehr G, Reznick R. Using operative outcome to assess technical skill. *The American Journal of Surgery 2000;180-3:234-7.*

109. Faulkner H. Validation of an objective structured assessment of technical skill for surgical residents. *Academic Medicine 1996;71-12:1363-65.*

110. Paisley AM, Baldwin PJ, Paterson-Brown S. Validity of surgical simulation for the assessment of operative skill. *British journal of surgery 2001;88-11.*

111. Cosman PH, Hugh TJ, Shearer CJC, Merrett NDC, Biankin AVC, Cartmill JA. Skills acquired on virtual reality laparoscopic simulators transfer into the operating room in a blinded, randomised, controlled trial. *Medicine Meets Virtual Reality 15: In Vivo, in Vitro, in Silico: Designing the Next in Medicine 2007:76.*

112. Madan AK, Frantzides CT. Prospective randomized controlled trial of laparoscopic trainers for basic laparoscopic skills acquisition. *Surgical endoscopy 2007;21-2:209-13.*

113. Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane database of systematic reviews (Online) 2009-1.*

114. Hu Y, Malthaner RA. The feasibility of three-dimensional displays of the thorax for preoperative planning in the surgical treatment of lung cancer. *European Journal of Cardio-Thoracic Surgery 2007;31-3:506-11.*

115. Starkes JL, Payl I, Jennen P, Leclair D. A stitch in time: cognitive issues in microsurgery. *Advances in psychology 1993;102:225-40.*

116. Starkes JL, Payk I, Hodges NJ. Developing a standardized test for the assessment of suturing skill in novice microsurgeons. *Microsurgery 1998;18-1.*

117. Praamsma M, Carnahan H, Backstein D, Veillette CJH, Gonzalez D, Dubrowski A. Drilling sounds are used by surgeons and intermediate residents, but not novice orthopedic trainees, to guide drilling motions. *Canadian Journal of Surgery 2008;51-6:442.*

118. Liu A, Tendick F, Cleary K, Kaufmann C. A survey of surgical simulation: applications, technology, and education. *Presence: Teleoperators & Virtual Environments 2003;12-6:599-614.*

119. Wei Q, Hu Y, MacGregor JH, Gelfand G. Segmentation of lung lobes in clinical CT images. *International Journal of Computer Assisted Radiology and Surgery 2008;3-1:151-63.*

120. Madan AK, Frantzides CT, Tebbit C, Quiros RM. Participants'opinions of laparoscopic training devices after a basic laparoscopic training course. *The American Journal of Surgery 2005;189-6:758-61.*

APPENDICES

APPENDIX 1: INFORMED CONSENT FORM

UNIVERSITY OF CALGARY

calgary health region

TITLE: The Feasibility and Fidelity of Practicing Surgical Fixation of an Ulna Fracture on Virtual Bone

INVESTIGATORS: Dr. Tyrone Donnon, Dr. Carol Hutchison, Dr. Justin LeBlanc

SPONSOR: Bone & Joint Health Research Portfolio

This consent form is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, please ask. Take the time to read this carefully and to understand any accompanying information. You will receive a copy of this form.

BACKGROUND

New learning techniques must be developed so that the knowledge and safety of the residents will be available when they enter the operating rooms. At this time, Sawbones and cadavers are more commonly used, but bring with them a great expense. A new method of learning is needed and virtual simulation has proved to be a valuable tool in many surgical fields. It is felt that virtual simulators can be used for training in the field of orthopaedic trauma. To assess this, a simple surgical fixation of a fractured ulna can be performed on a virtual simulator and compared to that on a Sawbones model. You will be one of the surgical residents involved in this study. If this virtual surgical simulator is proven to be valuable, it can then be further used to evaluate more complicated fracture fixation techniques. Ideally, this will be a useful method for hands-on teaching aiding surgical residents in their training, as well as being used for an evaluation tool for programs to ensure proper techniques are being used.

WHAT IS THE PURPOSE OF THE STUDY?

The objectives of the proposed study are:
To compare the fidelities of the virtual bone simulator and Sawbones simulator at all levels of residency training on the surgical fixation subtasks of an ulna fracture;
To measure the performance and quality of surgical skills using a previously constructed checklist format;

To evaluate the learners perceptions of their experiences related to specific subtasks of the procedure;
To determine if it is feasible to develop a virtual fracture fixation model for orthopaedic surgery residents to perform surgical fixation of an ulna

WHAT WOULD I HAVE TO DO?
As the resident participating in this study, you will be asked to perform surgical fixation of the ulna on both the sawbones model as well as the virtual simulator. These should each take approximately 15 minutes of your time. After each procedure has been completed you will be asked to fill out a questionnaire, which will take approximately 10 minutes of your time each. There will be no follow-up after this session, and you will be able to leave once the questionnaire is done. There will be someone available to you at all times for asking questions pertaining to the study.

WHAT ARE THE RISKS?
There are no added risks to you during fixation of the sawbones compared to surgical fixation in the operating room. You will be asked to use power tools and the ends of the drill bits and screws are sharp, so you will be asked to handle these with care. There are no risks associated with the use of the simulator. The results of these procedures will be kept confidential and will in no way be used as a formal evaluation.

ARE THERE ANY REPRODUCTIVE RISKS?

No

WILL I BENEFIT IF I TAKE PART?

As a participant to this study, you will benefit by practicing surgical fixation of the ulna on a sawbones model, and you will be the first to attempt this procedure on a virtual simulator. This may become an important aspect to surgical resident education and may become a worthy tool for future surgical education and evaluations.
The information we acquire from this study may help us to provide more useful methods of surgical education for residents. Using virtual simulators for practicing surgical techniques is important for ensuring the resident is prepared before entering the operating room, which will benefit both the patient and save costs in terms of operating time.

DO I HAVE TO PARTICIPATE?

Your participation in this study is completely voluntary and you may withdraw at anytime. As much as your volunteering is appreciated, the researcher may also withdraw you from the study if they feel it is needed.

WHAT ELSE DOES MY PARTICIPATION INVOLVE?

Before and after performing the surgical fixation of the ulna on both models, you will be asked to fill out questionnaires developed specifically for this project. They should each take 10 to 15 minutes of your time, and someone will be available for you to ask any questions you may have.

WILL I BE PAID FOR PARTICIPATING, OR DO I HAVE TO PAY FOR ANYTHING?

Your time and patience is very appreciated. We will reimburse you for your parking ticket on the main campus of the University of Calgary.

WILL MY RECORDS BE KEPT PRIVATE?

All data collected about you will have a unique numerical identifier. These will include the checklists during your procedures, the virtual simulator data, and the questionnaire. The questionnaires will be placed anonymously in a folder after completion. The resident in charge of the study will have access to information as a failsafe, however the identity of the subjects will not be disclosed to anyone else, besides the University of Calgary Conjoint Health Research Ethics Board.

IF I SUFFER A RESEARCH-RELATED INJURY, WILL I BE COMPENSATED?

In the event that you suffer injury as a result of participating in this research, no compensation will be provided to you by the Bone & Joint Health Research Portfolio, the University of Calgary, the Calgary Health Region or the Researchers. You still have all your legal rights. Nothing said in this consent form alters your right to seek damages.

SIGNATURES

Your signature on this form indicates that you have understood to your satisfaction the information regarding your participation in the research project and agree to participate as a subject. In no way does this waive your legal rights nor release the investigators, or involved institutions from their legal and professional responsibilities. You are free to withdraw from the study at any time without jeopardizing your health care. If you have further questions concerning matters related to this research, please contact:

Dr. Justin LeBlanc 403-797-1444 (pager 5829)

or

Dr. Tyrone Donnon 403-210-9682

If you have any questions concerning your rights as a possible participant in this research, please contact The Chair of the Conjoint Health Research Ethics Board at the Office of Medical Bioethics, 403-220-7990.

Participant's Name	**Signature and Date**
Investigator/Delegate's Name	**Signature and Date**
Witness' Name	**Signature and Date**

The University of Calgary Conjoint Health Research Ethics Board has approved this research study.

A signed copy of this consent form has been given to you to keep for your records and reference.

APPENDIX 2: PRE-PROCEDURE QUESTIONNAIRE

Pre-Procedure Questionnaire Part A. **ID #________**

1) What year of medical training are you in (circle)?

PGY-1 PGY-2 PGY-3 PGY-4 PGY-5

2) What is your sex (circle): Male Female

3) What is your age (circle): 20-24 25-29 30-34 35-40 >40

4) Are you right or left handed (please circle) Right Left

5) How many time have you performed >50% of the surgical fixation of an ulna? (circle)

0 1-2 3-5 6-9 10-13 14-16 >17

6) How many times have you practiced surgical fixation of a forearm bone using a Sawbones model? (circle)

0 1-3 4-6 7-9 ≥10

7) How many times have you used a virtual reality simulator for any reason (in or outside of the medical field)? (circle)

0 1-3 4-6 7-9 ≥10

8) How many times have you used a virtual reality simulation in the medical field? (circle)

0 1-3 4-6 7-9 ≥10

9) What is your previous video game experience? (circle all those that apply)

Never played Played as child Played within 5 years Still Playing

If still playing, how often (circle) Rare Monthly Weekly Daily

10) What is your preferred method of learning psychomotor skills? (Rank order from most "1" to least preferred "5")

Lecture _______ Small group _ Video _______ Demonstration________
Hands-on _____ Other: ____ (Specify: ______________________)

<u>Pre-Procedure Questionnaire Part B.</u>

Using 5-point Likert scale please answer following questions
(1 = inferior, 2 = poor, 3=average, 4 = good, 5 = excellent)

(Knowledge: General understanding of procedure) **(Skill: Skill level at performing procedure)** **(Comfort: Comfort level with procedure)**	**Inferior**	**Poor**	**Average**	**Good**	**Excellent**
1) My knowledge in **<u>management</u>** of an ulna fracture is:	1	2	3	4	5
2) My skill of the **<u>management</u>** of an ulna fracture is:	1	2	3	4	5
3) My comfort with **<u>managing</u>** an ulna fracture is:	1	2	3	4	5
4) My knowledge about **<u>performing</u>** surgical fixation of an ulna is:	1	2	3	4	5
5) My skill to **<u>perform</u>** surgical fixation of an ulna is:	1	2	3	4	5
6) My comfort in **<u>performing</u>** surgical fixation of an ulna **<u>alone</u>** is (no staff in room):	1	2	3	4	5
7) My comfort in **<u>performing</u>** surgical fixation of an ulna with **<u>staff</u>** is:	1	2	3	4	5
8) My knowledge of the tools indicated for fixation of an ulna is:	1	2	3	4	5
9) My skill to use the tools indicated for fixation of an ulna is:	1	2	3	4	5
10) My comfort with the tools indicated for fixation of an ulna is:	1	2	3	4	5

APPENDIX 3: SAWBONES SIMULATOR QUESTIONNAIRE

<u>Post-procedure (Sawbones) Questionnaire</u>

ID # ______

Type surgical simulator used first: (circle) Sawbones Virtual

Using 5-point Likert scale please answer following questions:
(1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

Sawbones Model Environment	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
1) The Sawbones model was responsive to the actions performed (overall)	**1**	**2**	**3**	**4**	**5**
2) The tools were not problematic to use for the Sawbones model	**1**	**2**	**3**	**4**	**5**
3) Visual representation of the forearm was realistic enough for the procedure	**1**	**2**	**3**	**4**	**5**
4) Visual representation of the tools in the Sawbones model are important in the performance of this procedure	**1**	**2**	**3**	**4**	**5**
5) The general performance using the Sawbones model was close in comparison to my general performance in the clinical settings	**1**	**2**	**3**	**4**	**5**

Using 5-point Likert scale please answer following questions: (1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree) **Sawbones Model Equipment**	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
6) The Sawbones model demonstrated precise movements of tools	**1**	**2**	**3**	**4**	**5**

7) All tools/equipment required were accessible during the Sawbones model simulation	**1**	**2**	**3**	**4**	**5**
8) Tactile force feedback was simulated accurately on the Sawbones model	**1**	**2**	**3**	**4**	**5**
9) Placement of tools was properly simulated on the Sawbones model	**1**	**2**	**3**	**4**	**5**
10) Drilling through bone was accurate on the Sawbones model	**1**	**2**	**3**	**4**	**5**
11) Plunging (exiting second cortex) of the drill was easy to feel on the Sawbones model	**1**	**2**	**3**	**4**	**5**

<u>Sawbones Model Psychological</u>	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
12) While performing this procedure on the Sawbones model, it felt like I was actually doing the procedure on a patient	**1**	**2**	**3**	**4**	**5**
13) I felt comfortable performing the procedure	**1**	**2**	**3**	**4**	**5**
15) I felt like all my senses were engaged during the procedure	**1**	**2**	**3**	**4**	**5**
14) The actual drilling made me feel as though I were performing a real procedure (in OR)	**1**	**2**	**3**	**4**	**5**
16) The visual aspects of the environment (i.e. Sawbones, tools, table) made me feel as if I were performing the real procedure (in OR)	**1**	**2**	**3**	**4**	**5**
17) The feel of the equipment made me feel as if I were actually doing the real procedure (in OR)	**1**	**2**	**3**	**4**	**5**
18) The events around me made me feel as though I were actually doing the real procedure (in OR)	**1**	**2**	**3**	**4**	**5**
19) My experience in the Sawbones environment (overall) seemed consistent with my real world experiences	**1**	**2**	**3**	**4**	**5**

Using 5-point Likert scale please answer following questions: (1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
20) The Sawbones model is an effective method for learning surgical fixation procedures	**1**	**2**	**3**	**4**	**5**
21) The Sawbones model is effective for the **introduction** of basic surgical skills	**1**	**2**	**3**	**4**	**5**
22) The Sawbones model is an effective method to practice **previously learned** techniques for my surgical training	**1**	**2**	**3**	**4**	**5**
23) Sawbones model based examinations would be useful for the assessment of surgical fixation of the ulna	**1**	**2**	**3**	**4**	**5**
24) The Sawbones model would be valuable for refresher skills	**1**	**2**	**3**	**4**	**5**
25) People were available to answer my questions when needed during the procedure	**1**	**2**	**3**	**4**	**5**
26) The Sawbones model (overall) provided a challenging surgical experience	**1**	**2**	**3**	**4**	**5**
27) Further development of the Sawbones model is needed prior to formal evaluation tool	**1**	**2**	**3**	**4**	**5**
28) Prior Sawbones course/experience is needed prior to examination using the Sawbones model	**1**	**2**	**3**	**4**	**5**
29) I would likely use a Sawbones model in my spare time for practicing procedures, if it were readily available	**1**	**2**	**3**	**4**	**5**
30) I would be more likely to use a Sawbones model in my spare time than the virtual reality model for practicing procedures, if both were readily available (**answer only if second procedure completed**)	**1**	**2**	**3**	**4**	**5**
31) Fracture fixation using the Sawbones model was a valuable experience	**1**	**2**	**3**	**4**	**5**
32) This Sawbones model should be included in residency training program	**1**	**2**	**3**	**4**	**5**
33) A surgical skills laboratory would be valuable to my surgical training	**1**	**2**	**3**	**4**	**5**

Short answers: (write on back if more space is needed)

What were the strengths and weakness of the Sawbones Surgical Simulator?

Advantages:

Disadvantages:

What were your frustrations with the Sawbones Simulator?

What would you change on the Sawbones simulator?

What do you see as being the benefits of using a Sawbones simulator:

Other Comments ? (please write on back)

Your participation was much appreciated
Thank you

APPENDIX 4: VIRTUAL SIMULATOR QUESTIONNAIRE

<u>Post-procedure (Virtual) Questionnaire Part A.</u>

ID#______

Type surgical simulator used first: (circle) Sawbones Virtual

Using 5-point Likert scale please answer following questions:
(1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

Virtual Simulator Environment	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1) The virtual simulation was responsive to the actions performed (overall)	**1**	**2**	**3**	**4**	**5**
2) The controls were not problematic to use in the virtual model	**1**	**2**	**3**	**4**	**5**
3) Visual representation of the virtual forearm was realistic enough for the procedure	**1**	**2**	**3**	**4**	**5**
4) Visual representation of the tools in the virtual model are important in the performance of this procedure	**1**	**2**	**3**	**4**	**5**
5) The general performance using the virtual simulator was close in comparison to my general performance in the clinical settings	**1**	**2**	**3**	**4**	**5**

Using 5-point Likert scale please answer following questions:
(1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

Virtual Simulator Equipment	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
6) The virtual simulator demonstrated precise movements of tools	**1**	**2**	**3**	**4**	**5**
7) All tools/equipment required were accessible during the virtual procedure	**1**	**2**	**3**	**4**	**5**
8) Tactile force feedback was simulated accurately on the virtual model	**1**	**2**	**3**	**4**	**5**
9) Placement of tools was properly simulated on the virtual model	**1**	**2**	**3**	**4**	**5**
10) Drilling through bone was accurate on the virtual simulator	**1**	**2**	**3**	**4**	**5**
11) Plunging (exiting second cortex) of the drill was easy to feel on the virtual model	**1**	**2**	**3**	**4**	**5**

Virtual Simulator Psychological	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
12) While performing this procedure on the Virtual simulator, it felt like I was actually doing the procedure on a patient	**1**	**2**	**3**	**4**	**5**
13) I felt comfortable performing the procedure	**1**	**2**	**3**	**4**	**5**
14) I felt like all my senses (not sound) were engaged during the procedure	**1**	**2**	**3**	**4**	**5**
15) The actual drilling made me feel as though I were performing a real procedure (in OR)	**1**	**2**	**3**	**4**	**5**
16) The visual aspects of the environment (i.e. Display, Haptics device, table) made me feel as if I were performing the real procedure (in OR)	**1**	**2**	**3**	**4**	**5**

17) The feel of the equipment made me feel as if I were actually doing the real procedure (in OR)	**1**	**2**	**3**	**4**	**5**
18) The events around me made me feel as though I were actually doing the real procedure (in OR)	**1**	**2**	**3**	**4**	**5**
19) My experience in the Virtual environment (overall) seemed consistent with my real world experiences	**1**	**2**	**3**	**4**	**5**

Using 5-point Likert scale please answer following questions:
(1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
20) This virtual simulator is an effective method for learning basic surgical fixation procedures	**1**	**2**	**3**	**4**	**5**
21) This virtual simulator is effective for the **introduction** of basic surgical skills	**1**	**2**	**3**	**4**	**5**
22) This virtual simulator is an effective method to practice **previously learned** techniques for my surgical training	**1**	**2**	**3**	**4**	**5**
23) Virtual simulator based examinations would be useful for the assessment of surgical fixation of the ulna	**1**	**2**	**3**	**4**	**5**
24) The Virtual simulator would be valuable for refresher skills	**1**	**2**	**3**	**4**	**5**
25) People were available to answer my questions when needed during the procedure	**1**	**2**	**3**	**4**	**5**
26) The virtual simulator (overall) provided a challenging surgical experience	**1**	**2**	**3**	**4**	**5**
27) Further development of the virtual simulator is needed prior to it being a formal evaluation tool	**1**	**2**	**3**	**4**	**5**
28) Prior simulation experience is needed prior to examination using the virtual model	**1**	**2**	**3**	**4**	**5**
29) I would likely use a virtual model in my spare time for practicing procedures, if it were readily available	**1**	**2**	**3**	**4**	**5**

30) I would be more likely to use a virtual model in my spare time than the Sawbones model for practicing procedures, if both were readily available **(answer only if second procedure completed)**	**1**	**2**	**3**	**4**	**5**
31) Fracture fixation using the virtual model was a valuable experience	**1**	**2**	**3**	**4**	**5**
32) This virtual simulator should be included in residency training program	**1**	**2**	**3**	**4**	**5**
33) A surgical skills laboratory would be valuable to my surgical training	**1**	**2**	**3**	**4**	**5**

Short answers: (write on back if more space is needed)

What were the strengths and weakness of the Virtual Surgical simulator?

Advantages:

____

Disadvantages:

____

What were your frustrations with the Virtual Simulator?

____

What would you change on the Virtual simulator?

____

What do you see as being the benefits of using a Virtual simulator:

____

Other Comments ? (please write on back)

____

Your participation was much appreciated!
Thank you

<u>Post-Procedure Questionnaire   Part B.</u>

Using 5-point Likert scale please answer following questions
(1 = Inferior, 2 = Poor, 3 = Average, 4 = Good, 5 = Excellent)

(Knowledge: General understanding of procedure) **(Skill: Skill level at performing procedure)** **(Comfort: Comfort level with procedure)**	**Inferior**	**Poor**	**Average**	**Good**	**Excellent**
1) My knowledge in **management** of an ulna fracture is:	1	2	3	4	5
2) My skill of the **management** of an ulna fracture is:	1	2	3	4	5
3) My comfort with **managing** an ulna fracture is:	1	2	3	4	5
4) My knowledge about **performing** surgical fixation of an ulna is:	1	2	3	4	5
5) My skill to **perform** surgical fixation of an ulna is:	1	2	3	4	5
6) My comfort in **performing** surgical fixation of an ulna **alone** is (no staff in room):	1	2	3	4	5
7) My comfort in **performing** surgical fixation of an ulna with **staff** is:	1	2	3	4	5
8) My knowledge of the tools indicated for fixation of an ulna is:	1	2	3	4	5
9) My skill to use the tools indicated for fixation of an ulna is:	1	2	3	4	5
10) My comfort with the tools indicated for fixation of an ulna is:	1	2	3	4	5

APPENDIX 5: TASK SPECIFIC CHECKLIST FOR SURGICAL FIXATION OF THE ULNA

Objective Checklist for Application of Neutralization Plate to the Ulna

Examiner initials _______ ID # ________

Circle simulator used: Sawbones Virtual Simulator

Please check the item if the candidate has performed the task correctly.

Application of Neutralization Plate	No (0)	Yes (1)
Fracture anatomically reduced	☐	☐
Template used to determine contour of plate	☐	☐
Plate centered over fracture (onto bone)	☐	☐
Place 2.5mm drill guide into plate hole closest to fracture line (proximal END)	☐	☐
Place 2.5mm drill bit into drill guide (in hole) and drill first screw hole	☐	☐
Drill did not plunge UNSAFELY for screw #1	☐	☐
Screw length measured with depth gauge	☐	☐
Place 3.5mm drill guide into drilled hole	☐	☐
Place 3.5 mm tap in 3.5 mm drill guide, and tap	☐	☐
Appropriate 3.5 mm cortex screw inserted into first hole	☐	☐
2.5mm drill guide placed in other hole nearest fracture (distal END)	☐	☐
Second screw inserted with correct steps	☐	☐
Drill did not plunge UNSAFELY for screw #2	☐	☐
All the remaining screws are inserted alternating from one side to the other, utilizing proper sequence of steps and tools	☐	☐
Drill did not plunge UNSAFELY for the remaining screws	☐	☐

Number of errors overall	
Time to completion (min:seconds)	

APPENDIX 6: GLOBAL RATING SCALE FOR SUGRICAL FIXATION FORTHE ULNA

Examiner initials_______ ID #_______

Circle simulator used: Sawbones Virtual Simulator

Please circle the number corresponding to the candidate's performance in each category.

Rating Key:

1	2	3	4	5
Inferior Below minimally accepted	**Poor** Minimally acceptable	**Average** Average/Acceptable	**Good** Superior level of skill	**Excellent** Expert top (10%)

Principles of fracture fixation

1	2	3	4	5
Poor Knowledge of Principles		Knows important concepts in this type of fracture fixation.		Knowledge of both basic & advanced principles

Definitive Fixation

1	2	3	4	5
Inappropriate fixation methods		Appropriate fixation methods. Stable fixation.		Achieves excellent fixation.

Flow of Operation

1	2	3	4	5
Frequently stopped operating. Unsteady and hesitant of equipment.		Demonstrated ability for forward planning with steady progression of operative procedure. Familiar with equipment		Well planned course of operation, effortless flow from one move to the next. Excellent knowledge of equipment.

Instrument Handling

1	2	3	4	5
Repeatedly makes tentative or awkward moves with instruments		Competent use of instruments although occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness

Time & Motion

1	2	3	4	5
Many unnecessary moves		Efficient time/motion but some unnecessary moves		Economy of movement and maximum efficiency

OVERALL PERFORMANCE

1	2	3	4	5
Inferior	**Poor**	**Average**	**Good**	**Excellent**